

AD-A193 558

ARI Research Note 88-13

Project A:
Improving the Selection, Classification and
Utilization of Army Enlisted Personnel

DTIC FILE COPY

**Literature Review:
Validity and Potential Usefulness of
Psychomotor Ability Tests for
Personnel Selection and Classification**

Jeffrey J. McHenry and Sharon R. Rose
Personnel Decisions Research Institute

FOR

CONTRACTING OFFICER'S REPRESENTATIVE
DEIRDRE KNAPP

SELECTION AND CLASSIFICATION TECHNICAL AREA
LAWRENCE M. HANSER, CHIEF

MANPOWER AND PERSONNEL RESEARCH LABORATORY
NEWELL K. EATON, DIRECTOR



U. S. Army

Research Institute for the Behavioral and Social Sciences

APRIL 1988

Approved for public release; distribution unlimited.



88429094

88 4 29 094

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

ADA 193558

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS --	
2a. SECURITY CLASSIFICATION AUTHORITY --		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --		5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-13	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --		7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences	
6a. NAME OF PERFORMING ORGANIZATION Human Resources Research Organization	6b. OFFICE SYMBOL (If applicable) HumRRO	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, Virginia 22333-5600	
6c. ADDRESS (City, State, and ZIP Code) 1100 South Washington Street Alexandria, Virginia 22314-4499		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA 903-82-C-0531	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION --	8b. OFFICE SYMBOL (If applicable) --	10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code) --		PROGRAM ELEMENT NO. 6.37.31.A	PROJECT NO. 2Q2637 31A792
		TASK NO. 2.3.2	WORK UNIT ACCESSION NO. 2.3.2.C1
11. TITLE (Include Security Classification) Literature Review: Validity and Potential Usefulness of Psychomotor Ability Tests for Personnel Selection and Classification			
12. PERSONAL AUTHOR(S) Jeffrey J. McHenry and Sharon R. Rose (Personnel Decisions Research Institute)			
13a. TYPE OF REPORT Literature Review	13b. TIME COVERED FROM Oct 82 TO 1984	14. DATE OF REPORT (Year, Month, Day) April 1988	15. PAGE COUNT 259
16. SUPPLEMENTARY NOTATION Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel (Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, U.S. Army Research Institute).			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
		First-Tour Evaluation, Job Performance, Literature Review, Personnel Classification, Personnel Selection, Predictors, Project A, Psychomotor Ability.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>→ The psychomotor ability literature was reviewed to determine the validity and potential usefulness of psychomotor ability tests for personnel selection and classification. Over 2,200 psychomotor test validity coefficients were located. These were tabulated by ability (using Fleishman's psychomotor ability taxonomy), criterion (e.g., school vs. training vs. job performance), and job type. Analyses of these data showed that psychomotor tests had been used successfully to predict training and job performance (i.e., $r_{xy} > .20$) for many different occupations. Barriers to the use of psychomotor tests were also investigated. Reliability data indicate that psychomotor measures are not unstable. Moreover, the possibility of using computerized tests in the future to assess psychomotor abilities should eradicate the problem of apparatus differences which has</p>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Lawrence M. Hanser		22b. TELEPHONE (Include Area Code) (202) 274-8275	22c. OFFICE SYMBOL PERI-RS

(OVER)

ARI RESEARCH NOTE 88-13

19. Abstract (continued)

historically plagued psychomotor testing. Data on the intercorrelations between psychomotor and cognitive-perceptual tests showed that there may be some overlap between spatial and mechanical ability tests and many psychomotor ability tests, but that this overlap is not great, (.20 < r < .30). The data also indicated that psychomotor abilities are almost totally uncorrelated with tests of general ability, g. Data on group differences were sparse, but suggested that group differences for psychomotor ability tests are generally less than those for cognitive-perceptual ability tests.

Taken together, these findings suggest a need for further psychomotor test development and validation research. A suggested priority for such research is provided.

Keywords: Personnel selection, Classification, Job performance.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel**

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
Colonel, IN
Commander

Research accomplished under contract
to the Department of the Army

Human Resources Research Organization

Technical review by:

Cleson Martin
Deirdre Knapp

This report, as submitted by the contractor, has been cleared for release to Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or other reference services such as the National Technical Information Service (NTIS). The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

PREFACE

This Research Note is one of three that present the results of a literature review conducted as part of Project A, a large-scale, multiyear research program intended to improve the selection and classification system for initial assignment of persons to U.S. Army Military Occupational Specialties. The research is sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences.

The three Research Notes each cover a separate domain of measures of human abilities, interests, and other attributes. Their titles are:

- o *Literature Review: Cognitive Abilities -- Theory, History, and Validity* by Jody L. Toquam, VyVy A. Corpe, Marvin D. Dunnette, and Margaret A. Keyes.
- o *Literature Review: Validity and Potential Usefulness of Psychomotor Ability Tests for Personnel Selection and Classification* by Jeffrey J. McHenry and Sharon R. Rose.
- o *Literature Review: Utility of Temperament, Biodata, and Interest Assessment for Predicting Job Performance* by Leaetta M. Hough (ed.).

The findings presented in these documents were used in the development of a battery of new tests and inventories for use in Project A. The focus of that development effort was to identify abilities and other human attributes that seemed "best bets" for predicting soldiers' job performance, and then to develop new measures for those attributes. These Research Notes, however, have usefulness beyond that particular applied problem. Many issues pertinent to the measurement and use of human abilities are described and discussed in each of these compilations.

The Research Notes describe the results and findings of the literature review, but do not describe the literature search process itself. Therefore, we provide a description of that process here.

The literature search was conducted by three research teams from the Personnel Decisions Research Institute. Each team was responsible for one of the three fairly broadly defined areas of human abilities or characteristics that are reported in the Research Notes: cognitive abilities; psychomotor abilities; and non-cognitive characteristics such as vocational interests, biographical data, and measures of temperament. While these domains were convenient for purposes of organizing and conducting literature search activities, they were not used as (nor intended to be) a final taxonomy of possible predictor measures.

The major part of the literature search was conducted in late 1982 and early 1983. Within each of the three areas, the teams carried out essentially the same steps:

1. Compile an exhaustive list of reports, articles, books, or other sources that were possibly relevant to Project A.

2. Review each item and determine its relevancy for the project's general purposes by examining the title and abstract (or other brief review).
3. Obtain the sources identified in the second step as being relevant.
4. For relevant materials, conduct a thorough review and transfer applicable information onto special review forms developed for the project.

In the first step, several activities were designed to insure that the list would be as comprehensive as possible. Several computerized searches of relevant data bases were performed. Across all three ability areas, more than 10,000 potential sources were identified via the computer searches. (Of course, many of these sources were identified as relevant in more than one area, and were thus counted more than once.)

In addition to the computerized searches, reference lists were obtained from recognized experts in each area, emphasizing the most recent research in the field. Several annotated bibliographies were obtained from military research laboratories. Finally, the last several years' editions of research journals that are frequently used in each ability area were scanned, as were more general sources such as textbooks, handbooks, and appropriate chapters in the *Annual Review of Psychology* (which reviews the most recent research in a number of conceptually distinct areas of psychology).

The majority of the items identified in the first steps proved not relevant to the applied purpose--that is, the identification and development of promising measures for personnel selection in the U.S. Army. These nonrelevant sources were weeded out in Step 2.

The relevant sources were obtained and reviewed, and team members completed two forms for each source: an Article Review form and a Predictor Review form (several of the latter could be prepared for each source). These forms were designed to capture, in a standard format, the essential information about the reviewed sources, which varied considerably in their organization and reporting styles.

The Article Review form contained eight sections: citation, abstract, list of predictors (keyed to the Predictor Review forms), description of criterion measures, description of sample(s), description of methodology, other results, and reviewer's comments. The Predictor Review form contained seven sections: description of predictor, reliability, norms/descriptive statistics, correlations with other predictors, correlations with criteria, adverse impact/differential validity/test fairness, and reviewer's recommendations (about the usefulness of the predictor). Each predictor was tentatively classified into an initial working taxonomy of predictor constructs.

The Review forms and the actual sources that had been located were used in two primary ways for Project A purposes. First, three working documents were written, one for each of the three areas. These working documents later evolved into the three Research Notes named above. These documents identified and summarized the literature with regard to issues important to the research being conducted, the most appropriate organization or taxonomy of the constructs in each area, and the validities of the various measures for different types of job performance criteria. Second, the predictors identified in the review were subjected to further, structured scrutiny in order to select tests and inventories for use in later activities of Project A.

As a set, the three Research Notes should provide a valuable resource for scientists, researchers, and personnel practitioners interested in measurement of individual differences in humans for various applied purposes, but especially for selection and classification.

VALIDITY AND POTENTIAL USEFULNESS OF PSYCHOMOTOR ABILITY TESTS FOR PERSONNEL SELECTION AND CLASSIFICATION: A REVIEW AND INTEGRATION OF THE LITERATURE

CONTENTS

	<u>Page</u>
SECTION I: INTRODUCTION	1
Background and Approach	1
Overview of Report	2
SECTION II: A TAXONOMY OF THE PSYCHOMOTOR ABILITY DOMAIN	4
Historical Overview	4
Fleishman's Factor Analytic Work	6
Criticisms of Fleishman's Psychomotor Ability Taxonomy	10
Factor Analysis vs. Molar Correlational Analysis	10
The Relationship between Factors and Abilities	17
Construct Explication of the Psychomotor Domain	22
A Final Word on the Taxonomy of the Psychomotor Domain	24
SECTION III: MOTOR SKILLS LEARNING	26
A Definition of Motor Skills Learning	26
Typical Motor Performance Learning Curves	26
The Course of Motor Skills Learning	28
Stages of Motor Skills Acquisition	28
The Role of Abilities in Skills Acquisition	29
Models of Skill-Ability Relationships	31
Additional Research on the Role of Individual Differences in Learning	32
Variables Affecting Motor Skills Learning	33
Massed vs. Distributed Practice	33
Part vs. Whole Task Learning	34
Stimulus-Response Compatibility	35
Knowledge of Results	37
Feedback	39
Theories of Motor Skills Learning	39
Habit Theory	39
Cybernetic Theories	39
Hierarchical Control Theories	41
Information Processing Theories	42
An Integrative Model of Motor Skills Learning	44
Implications of Motor Skills Learning Theories for Psychomotor Ability Research	45
SECTION IV: THE USE OF PSYCHOMOTOR TESTS IN APPLIED PSYCHOLOGICAL RESEARCH	47
Content Validation of Psychomotor Tests	47
The Criterion-Related Validity Literature Review Process	51
Identifying Relevant Articles, Reports, and Manuals	51
Tabulation of Validity Coefficients	52
Method of Summarizing Validity Information	54

	<u>Page</u>
The Criterion-Related Validity Evidence for Psychomotor Tests	63
Professional, Technical, and Managerial Jobs	63
Clerical Jobs	75
Sales Jobs	75
Protective Service Jobs	75
Service Jobs	80
Mechanical and Structural Maintenance Jobs	89
Electronics Jobs	94
Industrial Jobs	97
Miscellaneous Jobs	100
Validity Results by Research Setting	105
Validity Summary for All Jobs	113
Comparison of Criterion-Related Validities by Job Type and Criterion Category	113
Summary of the Validity Evidence	119
 SECTION V: CURRENT ISSUES IN RESEARCH ON PSYCHOMOTOR ABILITIES . .	 121
Stability and Differential Stability	121
Utility of Psychomotor Predictors	132
Correlations between Psychomotor and Cognitive- Perceptual Abilities	133
Psychomotor Test Administration	137
Psychomotor Testing and Group Differences	141
The GATB and Group Differences	142
Group Differences and the ACB	146
Group Differences and the Cross-Cultural Aircrew Aptitude Battery (CCAAB)	151
Summary of Group Differences Data	153
The Use of Single- vs. Multiple Construct Predictors	153
 SECTION VI: SUMMARY AND RECOMMENDATIONS	 161
Summary of Major Findings	161
A Suggested Priority for Psychomotor Test Development Efforts .	163
Multilimb Coordination	164
Control Precision	165
Manual Dexterity	166
Rate control	166
Speed of Arm Movement	167
Wrist-Finger Speed	168
Finger Dexterity	169
Arm-Hand Steadiness	169
Aiming	170
Job Content and the Selection of Psychomotor Tests	170
 REFERENCES	 173
 APPENDIX A. Descriptions of Selected Psychomotor Tests	 A-1
B. References to Articles, Manuals, and Technical Reports Containing Psychomotor Validity Data	B-1

TABLES

<u>Table</u>		<u>Page</u>
1	Psychomotor Ability Factors Identified by Fleishman and His Associates	11
2	Fleishman's Taxonomy of Psychomotor Abilities	12
3	One-Day Test-Retest Reliability of Information Processing Abilities	20
4	Percentage of Pilot Training Failures Attributable in Whole or in Part to Problems Involving Coordination and Technique	50
5	Criterion Constructs	55
6	Job Types and Sample Jobs	57
7	Psychomotor Tests Used in Criterion-Related Validity Research	59
8	Validity Summary for Military Professional, Technical, and Managerial Jobs	64
9	Validity Summary for All Pilots	67
10	Validity Summary for Nonmilitary Professional, Technical, and Managerial Jobs	69
11	Validity Summary for All Professional, Technical, and Managerial Jobs	71
12	Validity Summary for Nonmilitary Clerical Jobs	73
13	Validity Summary for Nonmilitary Sales Jobs	76
14	Validity Summary for Military Protective Service Jobs	78
15	Validity Summary for Nonmilitary Protective Service Jobs	81
16	Validity Summary for All Protective Service Jobs	83
17	Validity Summary for Military Service Jobs	85
18	Validity Summary for Nonmilitary Service Jobs	87
19	Validity Summary for All Service Jobs	90
20	Validity Summary for Nonmilitary Mechanical and Structural Maintenance Jobs	92
21	Validity Summary for Nonmilitary Electronics Jobs	95

TABLES (continued)

<u>Table</u>		<u>Page</u>
22	Validity Summary for Nonmilitary Industrial Jobs	98
23	Validity Summary for Military Miscellaneous Jobs	101
24	Validity Summary for Nonmilitary Miscellaneous Jobs	103
25	Validity Summary for All Miscellaneous Jobs	106
26	Validity Summary for All Military Jobs	108
27	Validity Summary for All Nonmilitary Jobs	110
28	Validity Summary for All Jobs	114
29	Median Validities by Job Type and Criterion Category	116
30	Reliabilities for Selected Psychomotor Tests	123
31	Reliability of PETER Psychomotor Tests	131
32	Correlations Between Psychomotor and Cognitive- Perceptual Abilities	135
33	Computerized Psychomotor Tests in Use in the Military	140
34	Percentage of Variance Attributable to Age and Socioeconomic Status for the Three GATB Psychomotor Scales	143
35	Ethnic Group Differences in GATB Psychomotor Aptitude Scores	145
36	Descriptive Statistics and Validities for Selected Psychomotor Tests in the Aircrew Classification Battery for Samples of Black and White Aviation Students	147
37	Descriptive Statistics and Validities for Selected Psychomotor Tests in the Aircrew Classification Battery for Samples of Male and Female Aviation Students	148
38	Validity of Three Psychomotor Tests for Predicting Graduation from Pilot Training for Foreign and American Student Pilots	152
39	Summary of Group Differences Data for Nine Psychomotor Abilities	154

FIGURES

<u>Figure</u>		<u>Page</u>
1	Typical learning curves for four basic motor skills learning research paradigms	27
2	Four-choice reaction time apparatus	36
3	Pathway apparatus testing reaction time.	37
4	An information processing model of motor skills learning	43

SECTION I

INTRODUCTION

Background and Approach

The psychomotor domain includes a broad class of abilities involved in coordinative, manipulative, repetitive and/or precise body or limb movements (Imhoff & Levine, 1981). The abilities represented in the psychomotor domain vary greatly in terms of speed, precision, and cognitive-perceptual requirements. All of the psychomotor abilities, however, involve motor movement. Indeed, as the domain name "psychomotor" suggests, motor movement is a defining characteristic of these abilities.

Given the perceptual-cognitive component involved in many psychomotor abilities, it should not be surprising to discover that the distinction between the cognitive-perceptual and psychomotor domains is not always very clear. For example, some of the abilities often included in the cognitive domain (e.g., reaction time) involve a very minor movement component. Similarly, many psychomotor abilities (e.g., rate control, control precision, and multilimb coordination) are highly dependent on cognitive-perceptual processes such as attention, encoding, comparing, and movement judgment. For purposes of this part of the report, we focus on only those abilities that involve a major manipulative or movement component.

This is quite similar to the distinction Imhoff and Levine drew between psychomotor and perceptual abilities in their 1981 literature review. Nevertheless, Imhoff and Levine chose not to retain this distinction throughout most of their paper. Instead, they combined the perceptual and psychomotor domains into a single perceptual-motor domain. This contrasts with the approach followed here.

The approach here also differs slightly from Fleishman's conceptualization of the limits of the psychomotor domain, which included reaction time and response orientation (i.e., choice reaction time) as two psychomotor ability dimensions. In our view, performance on measures of reaction time and response orientation is determined almost entirely by cognitive processes. During a typical assessment of these abilities, a subject's only motor task is to lift a finger from a button. This scarcely approaches the richness of psychomotor performance (i.e., coordinative, manipulative, repetitive, and/or precise body or limb movements) described above.

Time sharing (or divided attention) measures are very frequently assessed using psychomotor tasks (e.g., tracking tasks). To date, many researchers who have attempted to assess time sharing have totally confounded time sharing with the abilities underlying performance on the individual tasks used in the time sharing measure (e.g., Braune & Wickens, 1983; Damos, 1978; Owens, Goodman, Pollack, & Braune, 1983; cf. Ackerman, Schneider, & Wickens, 1982). Most psychologists, though, would agree with Damos (1978), who stated that the ability to time share is dependent on a subject's attentional-processing capacity. Therefore, time sharing ability is not included in the section on psychomotor abilities.

Overview of Report

The balance of this part of the report is devoted to a review of the literature on psychomotor abilities, with an emphasis on the use of psychomotor abilities as predictors of training and job performance.

First, the abilities represented in the psychomotor domain are identified and defined. In this section, there is also a brief history of psychomotor ability research. At the conclusion of this section, a taxonomy of psychomotor abilities is presented. This taxonomy is used in a subsequent presentation of the validity evidence for psychomotor abilities.

The next section contains a brief review and summary of research on motor skills learning and acquisition. The review of motor skills learning includes a description of common experimental paradigms, a summary of major research findings, and a listing of some of the parameters which are believed to govern motor skills learning. There is also a discussion of Fleishman's views regarding the differences between psychomotor skills and psychomotor abilities. The section concludes with a presentation of several competing theories of the motor skills learning process.

The use of psychomotor tests in applied psychological research is described in the following section. That section begins with a summary of the content validity evidence for psychomotor measures. Next, there is a summary of the criterion-related validity research using the same set of criterion constructs (i.e., educational and school achievement, training performance, job proficiency, and job involvement/withdrawal) used in the validity summary for the cognitive-perceptual domain. Separate validity summary tables are presented for nine different job types for both military and non-military jobs and subjects. In addition, because a large percentage of the military research has involved pilots and pilot trainees, a separate summary of these data is provided.

Following the presentation of the validity evidence, a section is devoted to four major issues in current research on psychomotor abilities: stability of psychomotor measures, utility, group differences, and the use of single- vs. multiple-construct predictor measures. The discussion of stability focuses primarily on the reliability of the rank ordering of individuals' psychomotor performance as psychomotor tasks are practiced and learned. The discussion of the utility of psychomotor tests covers issues such as administrative efficiency and cost-effectiveness. Different test administration formats are examined (e.g., paper-and-pencil measures, apparatus measures, and computerized measures). Also, the intercorrelations between psychomotor and cognitive-perceptual abilities are presented in order to determine the magnitude of the unique variance which psychomotor measures might contribute to the prediction of job performance and related criteria. With respect to group differences, the evidence is scant. There is a brief review of the few studies which compare the validity of psychomotor measures for blacks and whites. In addition, two validity studies conducted with samples of foreign student pilots are described. Finally in this section, there is an examination of the controversy surrounding multiple-construct psychomotor predictor measures and the practical utility of such measures in large-scale selection research.

The last section of this report contains a summary of the previous sections. The focus of this summary is the implications of the psychomotor literature for the ongoing Army selection and classification research project. Recommendations are made in light of the goals and objectives of this research effort.

SECTION II

A TAXONOMY OF THE PSYCHOMOTOR ABILITY DOMAIN

Historical Overview

Much of what we know about the taxonomy of the psychomotor ability domain is based on a series of studies of airmen and basic trainee airmen by Fleishman and his associates. Most of this research is actually quite recent. Even though psychologists have been administering psychomotor tests since the 1890s and have been using psychomotor tests to predict pilot training success and job performance since before World War I, it was not until 1953 that Fleishman began a systematic investigation of the structure of the abilities comprising the psychomotor domain. This is extremely slow progress in comparison with the progress made by psychologists interested in the taxonomy of cognitive-perceptual abilities (e.g., Spearman, 1927; Thurstone, 1938).

Passey and McLaurin (1966) have reviewed the literature on early aircrew selection. Their description of the research conducted prior to World War II suggests a lack of concern for either predictor or construct explication. On the criterion side, Passey and McLaurin report that there was very little research into what a pilot actually does on the job. It was not until 1941 that psychologists attempted any sort of a systematic job analysis for pilots (Miller, 1947). This basic job analytic work enabled Army Air Forces (AAF) psychologists to make tremendous strides in psychomotor predictor development during World War II. Whereas previous selection of psychomotor predictors had relied heavily on arm chair analysis concerning the appropriate psychomotor predictors for different jobs (cf. Viteles, 1942), AAF psychologists such as Melton (1947) and Miller now had the job analysis information needed to make a more objective determination of the psychomotor abilities and tests which might be relevant to pilot performance. Melton, in particular, was instrumental in developing a number of new apparatus tests. These tests, which became part of the Aircrew Classification Battery (ACB), were modified and improved frequently throughout the course of the war.

In the immediate postwar period, several events slowed the progress of psychomotor research. First, the end of the war greatly reduced the need for military manpower and personnel research. In addition, the separation of the Air Force from the Army fragmented the aggressive research efforts begun during the war.

Nevertheless, the Army, the Air Force, and the Navy continued to collect important data. These data included cognitive-perceptual measures as well as measures from psychomotor paper-and-pencil and apparatus tests. The accumulation of large data bases allowed researchers to use multivariate statistical techniques to identify the underlying factor structure of aircrew selection test batteries. Guilford and Lacey (1947), Dudek (1948, 1949), Michael (1949), and Roff (1951) carried out factor analyses based on data collected during the war, while Roff (1953) factor analyzed data collected immediately following the war in a joint Navy-Air Force study.

At the same time, a body of factor analytic studies of psychomotor test batteries was accumulating from nonmilitary research (e.g., Harrell,

1940; Seashore, Buxton, & McCollom, 1940; Wittenborn, 1945). There were a number of similarities between the factors identified in these nonmilitary studies and those identified in the military studies. The similarities were first summarized in a literature review by Fleishman (1953). Fleishman identified the following 10 psychomotor factors as prime candidates for future research:

1. Reaction Time. The speed with which a subject can make a simple, predetermined response upon presentation of a stimulus.
2. Tapping. The speed with which a subject can oscillate his fingers or his arm, independent of any eye-hand coordination.
3. Psychomotor Coordination. The integration of muscle movements and/or the coordination between eye and muscle movements.
4. Manual Dexterity. The ability to make skillful, speeded arm or hand movements.
5. Finger Dexterity. The ability to make skillful, fine, speeded object manipulations with the fingers.
6. Psychomotor Precision. An ambiguously defined ability which involves speeded, fine object manipulations with the fingers, but seemingly embraces more eye-hand coordination than finger dexterity.
7. Steadiness. The ability to make extremely coordinated, accurate movements in which the need for speed and strength are minimized.
8. Motor Kinesthesia. The ability to make a compensatory motor response in order to keep a unit balanced in a given position.
9. Aiming. The ability to execute a series of movements requiring eye-hand coordination.
10. Ambidexterity. The ability to perform rapid, fairly accurate movements with one's non-preferred hand.

In the sections that follow, the psychomotor ability factors above and additional psychomotor ability factors identified in subsequent research are discussed and described in terms of the psychomotor tests which load and define the ability factors. To facilitate understanding of these psychomotor tests, brief summaries of many of the most commonly used psychomotor tests are provided in Appendix A. The summaries include: the name of the test; the name and/or description of the ability construct which the test measures; a brief description of the task which the subject is required to perform for the test, generally accompanied by a picture of the apparatus for apparatus tests or a sample of test items for paper-and-pencil tests; a summary of administration and scoring instructions; and a summary of the reliability and validity evidence for the test.

Fleishman's Factor Analytic Work

During the mid and late 1950s, Fleishman and his associates continued their research into the taxonomy of the psychomotor ability domain. Fleishman (1967) has stated that the primary impetus for these investigations was to uncover the ability factors common to both psychomotor tests and pilot performance. Fleishman found it at least somewhat surprising that many of the valid psychomotor predictors of pilot performance bore little superficial resemblance to the tasks pilots performed on the job. A prime example was the Rotary Pursuit Test (see Appendix A). A subject's task in this test is to keep a metal stylus in contact with a small metallic target which is set on a rotating disk. The disk resembles a phonograph turntable in both appearance and operation. Fleishman (1954b) reported a validity of .27 for the Rotary Pursuit Test as a predictor of graduation from pilot training. This and other correlations between psychomotor tests and pilot performance indicated to Fleishman that there must be a common set of abilities underlying performance on both the tests and the pilot's job.

Fleishman was certainly not the first who sought to identify the abilities underlying psychomotor and pilot performance. As was noted above, before Melton, Miller, and their colleagues designed the Airman Classification Battery (ACB) psychomotor tests during World War II, they attempted to do a careful analysis of pilots' jobs in order to identify the abilities contributing to successful pilot performance. They then constructed their new psychomotor apparatus tests specifically to assess these hypothesized underlying abilities. Fleishman was simply extending the work of Melton and Miller by using factor analysis to identify and define these psychomotor abilities more precisely. In these efforts, Fleishman was greatly aided by the research which preceded his. For example, his 1953 review of previous factor analytic research cited ten previously suggested factors meriting further study. Fleishman also drew heavily upon the apparatus tests of the ACB (Melton, 1947) in the conduct of his research.

Between 1953 and 1962, Fleishman authored or co-authored numerous studies on the taxonomy of the psychomotor ability domain. In almost all of these studies, Fleishman administered a large battery of psychomotor tests (usually both apparatus and paper-and-pencil tests) to a sample of pilot or airmen trainees. In some of these studies, the psychomotor tests were supplemented by a battery of physical performance or cognitive-perceptual ability tests. Fleishman typically factor analyzed the correlation matrix for the test battery using Thurstone's (1947) centroid method. The resulting factor pattern matrix was then rotated to achieve simple structure. In the following paragraphs, eight such studies are summarized.

Fleishman and Hempel (1954a) administered a battery of 15 dexterity tests to a sample of 400 basic trainee airmen. These tests included a number of widely used apparatus tests (e.g., O'Connor Finger Dexterity, Purdue Pegboard, Minnesota Rate of Manipulation, and Santa Ana Finger Dexterity) as well as several paper-and-pencil tests designed to measure a subject's speed and accuracy in making marks inside figures and tracing lines (e.g., Large Tapping, Small Tapping, Tracing, and Square Marking). Fleishman and Hempel's goal was to identify the factors underlying manipulative performance. Via factor analysis, they were able to extract and identify five ability factors: finger dexterity, manual dexterity,

wrist-finger speed, aiming, and positioning. Three of these abilities--finger dexterity, manual dexterity, and aiming--corresponded exactly to factors Fleishman (1953) had identified in his literature review. A fourth factor, wrist-finger speed, was just a new name for the factor that Fleishman (1953) had previously called tapping. The fifth factor, positioning, was the least clearly defined. All of the tests loading on the positioning factor require subjects to place or position blocks, pegs, or pins into snugly fitting holes in an apparatus board as quickly as possible. Fleishman and Hempel (1954a) noted that this was somewhat similar to aiming, which involves making rapid, precise pencil marks inside a series of small circles. Still, they were unable to specify precisely the unique ability represented by this factor.

In a much larger study, Fleishman (1954a) administered 27 apparatus tests and 11 printed tests to a sample of 400 basic trainee airmen. Fleishman's goal was "to verify empirically in a single study practically all of the psychomotor factors previously identified in the separate studies, together with any new factors that might emerge" (p. 438). The test battery that Fleishman used included a number of AAF apparatus tests used during World War II (Melton, 1947) as well as the dexterity tests used in the Fleishman and Hempel (1954a) study cited above. Of the nine well-defined factors which emerged from the factor analysis, seven had been identified previously: wrist-finger speed, finger dexterity, aiming, arm-hand steadiness, reaction time, manual dexterity, and psychomotor coordination. The remaining two factors Fleishman called rate of arm movement and spatial relations. The tests defining rate of arm movement included Ten Target Aiming, Two-Plate Tapping, and Rotary Aiming. All three tests require subjects to make gross, rapid arm movements. The importance of accuracy is reduced or minimized in all of these tests. Tests loading the spatial relations factor included Discrimination Reaction Time and Complex Coordination. Both of these tests require subjects to determine the appropriate spatial direction for their responses in accordance with the stimulus presented.

Hempel and Fleishman (1955) subsequently factor analyzed a battery of 46 manipulative, paper-and-pencil, and physical performance tests administered to a sample of 400 basic trainee airmen. The 17 manipulative tests included a number of placement, assembly, and turning tasks. Most of the six printed tests required the subject to draw lines or place dots or X's inside figures. The remaining 23 physical performance tests were assessments of physical fitness (e.g., number of chin-ups completed in a fixed period of time). Hempel and Fleishman failed to indicate why they attempted to factor analyze such a mixed lot of tests. Not surprisingly, the physical performance tests did not load on the same factors as the manipulative and paper-and-pencil tests. Of the 14 interpretable factors, only four were relevant to the psychomotor domain: aiming, arm-hand steadiness, manual dexterity, and finger dexterity. All four of these ability factors had emerged in previous factor analytic studies.

A battery of 11 apparatus tests and nine paper-and-pencil tests was administered by Fleishman and Hempel (1955) to a sample of 264 basic trainee airmen. The apparatus tests included a discrimination reaction time test, four simple reaction time tests, two dexterity tests, two complex psychomotor tests taken from the ACB, and two speed of arm movement tests. The nine printed tests included a vocabulary test, a current affairs test,

three tests related to mechanical knowledge, and four tests of perceptual-visual ability. Nine factors emerged from a factor analysis of this test battery. Fleishman and Hempel regarded six of these as psychomotor ability factors: discrimination reaction time, reaction time, psychomotor coordination, rate of arm movement, finger dexterity, and spatial relations. The discrimination reaction time factor was test-specific; the only variables loading on this factor were scores taken from the Discrimination Reaction Time Test. The remaining five factors have all been identified and described previously.

In 1947-1948, over 1,000 Navy pilot candidates completed 23 paper-and-pencil and apparatus psychomotor tests as part of a joint Navy-Air Force research project noted briefly above (Roff, 1953). The 16 apparatus tests were drawn primarily from the tests used by the AAF during World War II, while the seven printed tests were intended to measure marking speed and visualization abilities. Fleishman and Hempel (1956) obtained the data for these subjects and submitted the data to a factor analysis. The results of the factor analysis were particularly interesting because this was the largest and most diverse set of psychomotor apparatus tests which had ever been included in a factor analysis. Nine factors emerged: two psychomotor coordination factors, two spatial relations factors, integration, rate control, perceptual speed, manual dexterity, and visualization.

Fleishman and Hempel (1956) indicated that the first psychomotor coordination factor was the ability to perform "muscular movements involved in making fine, accurate (control) adjustments" (p. 100). The tests which loaded most highly on this factor included Complex Coordination, Pursuit Confusion, Rotary Pursuit, Two-Hand Coordination, and Rudder Control. The tests loading most highly on the second psychomotor coordination factor were Rudder Control, Plane Control, Multidimensional Pursuit, Complex Coordination, and Two-Hand Coordination. All of these tests involve the ability to coordinate the movement of two limbs.

The tests loading on the first spatial relations factor all use complex stimulus patterns to cue the subject. The subject is then required to interpret some spatial characteristic of the stimulus pattern (e.g., the subject might have to judge his body orientation in relation to the orientation of the stimulus). After interpreting this spatial characteristic, the subject must make an appropriate response. According to Fleishman and Hempel, the most difficult aspect of this task is the interpretation of the stimulus pattern. By comparison, the determination and execution of the appropriate response are relatively simple. The tests loading on this first spatial relations factor included Controls Orientation, Drift Correction, Direction Control, and Directional Control. This first spatial relations factor contrasted sharply with the second spatial relations factor, which Fleishman and Hempel called response orientation. The second spatial relations factor was quite similar to the spatial relations factor which Fleishman and his colleagues had identified in two previous studies (Fleishman, 1954a; Fleishman & Hempel, 1955). The tests loading on this factor included Signal Discrimination, Discrimination Reaction Time, and Complex Multiple Reaction. All of these tests require subjects to quickly choose and execute the correct response upon presentation of the stimulus. Generally, for these tests the subject is presented with a rather simple stimulus (e.g., a single light). The subject is then required to determine the appropriate response from among a number of response alternatives. For

some tests, there is a complex spatial relationship between the location of the stimulus and the location or direction of the appropriate response. For other tests, a complex series of spatially related responses are required upon presentation of the stimulus. Thus, while the focus of complexity for the first spatial relations factor was on the stimulus, the focus of complexity for the second spatial relations factor was on the response.

The other two psychomotor abilities which were identified in this study were rate control and manual dexterity. The tests loading on the rate control factor included Rate Control, Single-Dimension Pursuitemeter, and Compensatory Balance. These tests all require subjects to make anticipatory control adjustments in response to changes in the speed and direction of a continuously moving stimulus. The manual dexterity factor had been identified and defined by Fleishman and his associates in a number of previous studies.

The remaining three factors identified by Fleishman and Hempel-- integration, perceptual speed, and visualization--were cognitive-perceptual ability factors.

This factor analytic study was particularly important because it suggested several new psychomotor ability factors. First, two psychomotor coordination factors emerged from the analysis. The first was concerned with highly controlled movements and adjustments, while the second focused on the ability to coordinate the movements of two or more limbs. A new, tentative spatial relations ability factor was also identified. The focus of this spatial relations factor was identifying and interpreting a spatial characteristic embedded within a complex stimulus pattern. In addition, the focus of the second spatial relations factor was more clearly specified as being the ability to choose and execute the correct response alternative upon presentation of a relatively simple stimulus. Finally, a rate control ability factor emerged. This factor was defined by tests requiring the subject to make continuous anticipatory motor judgments in response to a moving stimulus.

In a follow-up to this study, Fleishman (1958) administered a battery of 31 psychomotor apparatus tests to a sample of 204 basic trainee airmen. Seven interpretable factors emerged from the factor analysis: response orientation (corresponding to the second spatial relations factor from Fleishman & Hempel, 1956), fine control sensitivity (corresponding to the first psychomotor coordination factor from Fleishman & Hempel), reaction time, speed of arm movement (previously called rate of arm movement), arm-hand steadiness, multilimb coordination (corresponding to the second psychomotor coordination factor from Fleishman & Hempel), and rate control. This study helped confirm many of the psychomotor ability factors first identified by Fleishman and Hempel.

A study by Parker and Fleishman (1960) further confirmed the taxonomy of psychomotor ability factors identified by Fleishman and Hempel (1956) and Fleishman (1958). A total of 203 Air Force ROTC students at a large eastern university completed 29 psychomotor apparatus tests and 21 psychomotor and perceptual paper-and-pencil tests. Eight of the 15 factors which emerged from a factor analysis were defined primarily by psychomotor tests: control precision (corresponding to the first psychomotor coordination

factor from Fleishman & Hempel, and the fine control sensitivity factor from Fleishman, 1958), speed of arm movement, manual dexterity, reaction time, response orientation, arm-hand steadiness, finger dexterity, and multilimb coordination.

In the most recent factor analytic study of the psychomotor domain by Fleishman and his associates, Fleishman and Ellison (1962) administered a battery of 22 fine manipulative and dexterity tests to 760 airmen who were entering one of three technical schools: engine mechanic, hydraulic mechanic, and aircraft electrician. The battery included nine paper-and-pencil tests and 13 apparatus tests. Five interpretable factors emerged from a factor analysis: wrist-finger speed, manual dexterity, finger dexterity, aiming, and speed of arm movement. All five of these factors had emerged in at least one previous study (Fleishman, 1954a).

Table 1 provides a list of the psychomotor ability factors which emerged from each of the eight factor analytic studies summarized above. Of the 13 factors listed, all but the positioning factor and the "stimulus" spatial relations factor (called simply Spatial Relations in Table 1) were identified in at least two studies. Fleishman (1967) maintained that this was strong evidence for a psychomotor ability taxonomy consisting of 11 ability factors. Table 2 contains a list of these 11 abilities. Table 2 also lists marker tests for each ability (Appendix A contains a description of most of these marker tests) and provides references for some of the factor analytic studies in which the ability has been identified.

Further evidence for the validity of portions of this taxonomy is provided by a study of pilot performance (Fleishman & Ornstein, 1960). For this study, 63 graduates of an Air Force pilot school were required to perform a series of 24 flight maneuvers. These 24 maneuvers had been chosen to be representative of the total range of nonacrobatic maneuvers taught during pilot training. For each maneuver, a Daily Progress Record Sheet (DPRS) had been prepared (Sutter, Townsend, & Ornstein, 1954). The DPRS for each maneuver contained a list of items which had been judged to be essential to proper execution of that maneuver. As each pilot executed each maneuver, a trained rater marked the DPRS to indicate any errors made by the pilot. The pilot's score on the maneuver was the total number of errors he made on the individual items. Each pilot was rated four times on each maneuver. Fleishman and Ornstein subjected the error scores to a factor analysis in order to identify the factors underlying pilot performance. Six factors emerged from the analysis. Four of these performance factors paralleled ability factors previously identified by Fleishman and his associates: control precision, multilimb coordination, response orientation, and rate control. The other two factors, spatial orientation and kinesthetic discrimination, were similar to previously identified perceptual factors.

Criticisms of Fleishman's Psychomotor Ability Taxonomy

Fleishman's taxonomy of psychomotor abilities has been criticized by a number of researchers over the past 20 years. The criticisms center mainly on his use of factor analysis to identify ability dimensions.

Factor Analysis vs. Molar Correlational Analysis. Jones (1960, 1962) has been highly critical of the use of factor analysis to identify factors

Table 1

Psychomotor Ability Factors Identified by Fleishman and His Associates

Psychomotor Ability Factor	Study							
	Fleishman & Hempel, 1954a (N=400)	Fleishman, 1954a (N=400)	Hempel & Fleishman, 1955 (N=400)	Fleishman & Hempel, 1955 (N=264)	Fleishman & Hempel, 1956 (N=1000+)	Fleishman, 1958 (N=204)	Parker & Fleishman, 1960 (N=203)	Fleishman & Ellison, 1962 (N=760)
Multilimb Coordination		x ^b		x ^b	x ^e	x	x	
Control Precision		x ^b		x ^b	x ^d	x ^h	x	
Rate Control					x	x		
Finger Dexterity	x	x	x	x			x	x
Manual Dexterity	x	x	x		x	x	x	x
Wrist-Finger Speed	x	x	x					x
Aiming	x	x	x					x
Arm-Hand Steadiness		x				x	x	
Speed of Arm Movement		x ^a		x ^a			x	x
Reaction Time		x		x		x	x	
Response Orientation		x ^c		x ^c	x ^g	x	x	x
Positioning	x							
Spatial Relations								x ^f

a. In these articles, this factor was called Rate of Arm Movement.

b. These two abilities were both subsumed in a factor called Psychomotor Coordination.

c. In these articles, this factor was called Spatial Relations. In both studies, however, the tests loading on this factor are actually tests of Response Orientation.

d. In this article, this factor was called Psychomotor Coordination I.

e. In this article, this factor was called Psychomotor Coordination II.

f. In this article, this factor was called Spatial Relations I.

g. In this article, this factor was called Spatial Relations II.

h. In this article, this factor was called Fine Control Sensitivity.

Table 2

Fleishman's Taxonomy of Psychomotor Abilities

Psychomotor Ability Factor	Definition	Marker Tests	References
Multilimb Coordination	The ability to coordinate the simultaneous movement of two or more limbs. This ability is general to tasks requiring coordination of any two limbs (e.g., two hands, two feet, one foot and one hand). The ability does <u>not</u> apply to tasks in which trunk movements must be integrated with limb movements. It is most common to tasks where the body is at rest (e.g., seated or standing) while two or more limbs are in motion.	Complex Coordination Test Rudder Control Test	Fleishman, 1954a Fleishman & Hempel, 1955 Fleishman & Hempel, 1956 Fleishman, 1958 Parker & Fleishman, 1960
Control Precision	The ability to make fine, highly controlled (but not overcontrolled) muscular movements necessary to adjust or position a machine control mechanism. This ability is general to tasks requiring motor adjustments or movements in response to a stimulus whose speed and direction of movement are perfectly predictable. This ability is crucial in situations where the motor adjustments must be both rapid and precise. The ability extends to arm-hand movements as well as to leg movements.	Rotary Pursuit Test Control Adjustment Test	Fleishman, 1954a Fleishman & Hempel, 1955 Fleishman & Hempel, 1956 Fleishman, 1958 Parker & Fleishman, 1960
Rate Control	The ability to make continuous anticipatory muscular movements necessary to adjust or position a machine or equipment control mechanism. This ability is general to tasks requiring motor adjustments in response to a moving stimulus which is changing speed and/or direction in a random or unpredictable manner. The ability applies to compensatory tracking as well as following pursuit of the stimulus. This ability does <u>not</u> extend to situations in which both the speed and direction of the stimulus are perfectly predictable.	Motor Judgment Test Single Dimension Pursuitemeter Rate Control Test	Fleishman & Hempel, 1956 Fleishman, 1958

(Continued)

Table 2 (Continued)

Fleishman's Taxonomy of Psychomotor Abilities

Psychomotor Ability Factor	Definition	Marker Tests	References
Arm-Hand Steadiness	The ability to make precise, steady arm-hand positioning movements where both strength and speed are minimized. This ability includes steadiness during movement as well as minimization of tremor and drift while maintaining a static arm position. This ability does <u>not</u> extend to manipulation of machine or equipment control mechanisms.	Track Tracing Test Steadiness-Precision Test Arm-Hand Steadiness Test	Fleishman, 1954a Fleishman, 1958 Parker & Fleishman, 1960
Speed of Arm Movement	This ability involves the speed with which discrete arm movements can be made. The ability deals with the speed with which the movement can be carried out after it has been initiated; it is not concerned with the speed of initiation of the movement. Nor are the precision, accuracy, or coordination of the movement considered under this ability.	Movement Time (from a reaction time apparatus) Two-Plate Tapping	Fleishman, 1954a Fleishman & Hempel, 1955 Parker & Fleishman, 1960 Fleishman & Ellison, 1962
Reaction Time	The ability to perceive and react to a stimulus quickly.	Reaction Time	Fleishman, 1954a Fleishman & Hempel, 1955 Fleishman, 1958 Parker & Fleishman, 1960
Response Orientation	The ability to quickly choose the correct response to a stimulus from among two or more response alternatives.	Choice Reaction Time Discrimination Reaction Time Test	Fleishman, 1954a Fleishman & Hempel, 1955 Fleishman & Hempel, 1956 Fleishman, 1958 Parker & Fleishman, 1960

Table 2 (Continued)

Fleishman's Taxonomy of Psychomotor Abilities

Psychomotor Ability Factor	Definition	Marker Tests	References
Finger Dexterity	The ability to make skillful, coordinated, highly controlled movements of the fingers. This ability applies primarily to tasks involving manipulation of objects with the fingers. The ability does <u>not</u> extend to manipulation of machine or equipment control mechanisms.	Purdue Pegboard Test O'Connor Tweezer and Finger Dexterity Test General Aptitude Test Battery (GATB), Finger Dexterity Scale	Fleishman & Mempel, 1954a Fleishman, 1954a Mempel & Fleishman, 1955 Fleishman & Mempel, 1955 Parker & Fleishman, 1960 Fleishman & Ellison, 1962
Manual Dexterity	The ability to make skillful, coordinated movements of the hand or the arm and hand. This ability typically applies to tasks involving manipulation of moderately large objects (e.g., blocks, pencils, etc.) under speeded conditions. The ability does <u>not</u> extend to manipulation of machine or equipment control mechanisms.	Minnesota Rate of Manipulation Test General Aptitude Test Battery (GATB), Manual Dexterity Scale	Fleishman & Mempel, 1954a Fleishman, 1954a Mempel & Fleishman, 1955 Fleishman & Mempel, 1955 Fleishman & Mempel, 1956 Fleishman, 1958 Parker & Fleishman, 1960 Fleishman & Ellison, 1962
Wrist-Finger Speed	The ability to carry out very rapid, discrete movements of the fingers, hands, and wrists. This ability applies primarily to tasks in which the accuracy of the movement is <u>not</u> a major concern. The ability is <u>not</u> related to the speed of initiation of the movement, a la reaction time; it is determined entirely by the speed with which the movement is carried out.	General Aptitude Test Battery (GATB), Motor Coordination Scale Large Tapping Test	Fleishman & Mempel, 1954a Fleishman, 1954a Mempel & Fleishman, 1955 Fleishman & Ellison, 1962
Aiming	The ability to make very precise, accurate movements under highly speeded conditions. This ability is dependent upon very precise eye-hand coordination.	Small Tapping Test Trace Tapping II	Fleishman & Mempel, 1954a Fleishman, 1954a Mempel & Fleishman, 1955

(Continued)

underlying the pattern of correlations in a correlation matrix. According to Jones, factor analysis is an inductive procedure. Factor analysis permits no tests of theories or hypotheses. Rather, psychologists tend to execute factor analyses in a mechanical, unthinking fashion, according to Jones. They enter the raw data into a computer, compute the correlations among the variables, extract and rotate factors according to their favorite extraction and rotation algorithms, and then examine the resulting factor pattern matrix to identify the underlying factors.

Jones objects to this procedure on two grounds. First, he claims that factor analysis does not allow researchers to test a priori theories regarding the underlying factor structure of a set of variables. Jones prefers a more deductive, theory-testing approach to psychology. Second, Jones believes that one can learn a great deal about the factor structure of a set of variables simply by examining the pattern of correlations in a correlation matrix.

Based on his investigations, Jones (1960, 1962) concluded that the pattern of correlations within a correlation matrix typically assumes one of a relatively limited number of common patterns (e.g., simplex form, a circumplex form, monotonic hierarchy, rippled patterns, tiered patterns, superdiagonal form, etc.). These patterns are important because they tell a lot about the relationships between and among factors and variables. The patterns may suggest or reveal cumulative, competitive, developmental, and/or maturational relationships, for instance. This information is largely lost to the factor analyst, according to Jones.

As an alternative to factor analysis, Jones recommends that researchers begin by developing a theory about the factor structure underlying their set of variables. Hypotheses regarding the factor structure would likely focus on the number and possible nature of the factors, the particular sources of variance (i.e., common factors) shared by subsets of variables, and the particular group of factors comprising each variable. Theories could then be altered as needed according to the observed pattern of correlations in the correlation matrix. Jones suggests techniques for computing the factor pattern matrix for each of the common patterns of correlations he has identified. While the factor pattern matrix will typically not be an exact representation of the correlations in the correlation matrix, Jones (1960) states that one should be able to work backwards and use the factor pattern matrix to reproduce closely the observed correlations in the correlation matrix. For example, Jones's open contiguity model was such an accurate depiction of one correlation matrix that no observed correlation differed by more than .03 from the correlations reproduced from the factor pattern matrix.

Evaluating the utility of molar correlational analysis is somewhat difficult. Jones (1962) was correct to criticize Fleishman for relying solely on inductive methods of identifying factors. In practice, however, Jones's approach also relies quite heavily upon inductive methods. In almost all of the examples in his 1960 monograph, Jones proposed a model to explain the interrelationships among a set of variables only after he examined the correlation matrix. This is not unlike the approach of most factor analysts, who also rely on the pattern of correlations to determine the underlying factor structure of their variables. One might even say that factor analysts are a bit more objective in their approach to

identifying the factor structure. Factor analysts allow the computer to "choose" a structure for them based on a predetermined set of decision rules. Molar analysts, on the other hand, subjectively choose a model based on their perception of the pattern of correlations in the correlation matrix. There would seem to be more margin for error and unreliability with such an approach.

This is not to say that there are no potential advantages to Jones's methods. At least two are readily apparent.

First, molar correlational analysis would indeed permit researchers to test their theories regarding the factor structure of a set of variables. All researchers would need to do would be to specify *a priori* the structure of their data using one of Jones's models and then determine how well the computed factor pattern matrix accounted for the observed correlations in the correlation matrix. Unfortunately, molar correlational analysis provides no goodness-of-fit statistics for testing the correspondence between a data set and a theory, and Jones offers only minimal guidance for determining whether or not a factor model accurately depicts the observed data. Further, there have been no studies directed at the question of whether two or more different factor models might be able to account equally well for the same data set. No one has shown, for example, that the observed correlations in a correlation matrix in superdiagonal form could not be reproduced equally well by both an open contiguity structure and a simplex (two of the models described by Jones).

A second apparent advantage of molar correlational analysis is that it suggests the need to look for more complex relationships among variables and factors than factor analysts are typically likely to do. Most factor analysts continue to pursue Thurstone's (1947) goal of simple structure (i.e., all variables load on at most two or three factors and all factors are defined by a limited subset of variables). Jones presents a number of interesting examples where the principles of simple structure simply are not appropriate for the data under consideration. For example, he demonstrates that in a typical study involving practice, performance on early trials is determined by a number of factors whereas performance on later trials might be determined by only one or two factors (Jones, 1962). While this does not apply directly to Fleishman's taxonomic research, it does suggest instances where traditional factor analysis and rotation to simple structure are inappropriate. Stated simply, one must consider carefully the likely interrelationships among the factors and variables and choose the method of data analysis which is most appropriate for testing that hypothesized state.

An important reason for discounting many of Jones's criticisms, however, is that there have been a number of advances in structural modeling and confirmatory factor analysis over the past 20 years (e.g., Jöreskog, 1978) which would seem to be better suited than molar correlational analysis for answering the structural questions which Jones was posing. It is perhaps unfair to criticize Jones for failing to recommend statistical methodologies which were not to be explicated and popularized until 10-15 years after his 1960 monograph. Yet, critics of Fleishman often cite Jones's reanalysis of Fleishman and Hempel's (1954b) data in criticizing Fleishman's psychomotor taxonomy. Instead, the consistency in Fleishman's results (see Table 1) suggests that Fleishman's taxonomy would fare quite

favorably if his data were subjected to a confirmatory factor analysis.

The Relationship between Factors and Abilities. The second major criticism of Fleishman's psychomotor ability taxonomy centers on Fleishman's interpretation of the nature of the factors emerging from his analyses.

Fleishman (1975) has maintained that factor analysis is an appropriate method of identifying the abilities underlying performance on a series of tests or tasks. According to Fleishman, if two tests are highly correlated, it must be because a common ability (or set of abilities) underlies performance on both tests. Therefore, the factors emerging from a factor analysis of psychomotor tests can be interpreted by determining the ability common to performance on the tests loading significantly on each factor. Conversely, each factor must represent some real, underlying ability.

This view contrasts sharply with the opinions of several other trait researchers. For example, Eysenck has defined a factor as "a *hypothetical* [emphasis added] causal influence underlying and determining the observed relationships between a set of variables" (Eysenck, 1953, p. 108). According to Eysenck, factors are useful for providing a unifying structure for viewing the world. Factors are not necessarily real entities. Rather, as with any scientific law or postulate, factors are merely useful devices for making sense of nature (Thurstone, 1947). This position is similar to Anastasi's (1983). In her view, the primary purpose of factors is to provide a classification scheme for a set of variables. The basis of this classification scheme is, of course, the intercorrelations among the tests or variables.

To understand the nature of a factor, then, one must understand (or at least postulate) the common element or elements linking the tests or variables comprising the factor. Fleishman's position was that the common element was a real, underlying ability. Therefore, based on his factor analyses, he claimed that he was able to identify a psychomotor ability taxonomy. Anastasi (1983), however, claims that there are other reasons that tests or variables might be correlated. One of these seems particularly germane to Fleishman's analyses.

Anastasi states that one reason a factor might emerge is because of the contiguity or co-occurrence of learning experiences. For example, one of the factors which emerged in the Ohio State studies of leadership was a consideration factor (Halpin & Winer, 1957). This factor was comprised of items indicative of warmth, trust, friendship, interpersonal support, and mutual respect. In current psychological jargon, we would probably refer to this factor as interpersonal skill. Note, however, that even though this skill emerged as a unitary factor from a factor analysis, it is actually comprised of several distinct elements. For example, the ability to develop warm personal relations with others differs from the ability to earn others' respect. These two abilities would seem to rely on slightly different cognitive processes and would certainly be manifested behaviorally in quite different fashions. Yet, both of these are elements of the interpersonal skill factor. The most likely explanation of this unitary factor is that contingencies on interpersonal behavior operate at a rather general level. In any given social situation, an individual is likely to

have to draw on a number of the abilities comprising interpersonal skill.

As an individual's behavior in interpersonal situations is rewarded, punished, and shaped--depending in large measure upon the individual's current standing on the various component interpersonal abilities--a more general trait or ability may emerge. Individuals with generally high levels of interpersonal abilities may, for example, be rewarded for their interpersonal behavior. These rewards may provide them with the opportunities and confidence to practice and improve the interpersonal abilities in which they are less competent. Conversely, individuals who are adept at only a very few interpersonal behaviors may garner little or no reinforcement from their attempts to interact with others. As they withdraw from interpersonal situations, their few existing interpersonal abilities may "atrophy" due to lack of practice. Thus, based on this all-or-nothing reinforcement pattern, a number of somewhat disparate abilities may merge and appear as a broader, more general factor.

It is possible that the co-occurrence of learning experiences could help explain some of the factors emerging from Fleishman's analyses. For example, performance on measures of rate control (e.g., Motor Judgment Test, Rate Control Test) seems to depend on a number of abilities, including spatial orientation, timing, arm-hand steadiness, and hand-wrist motor coordination skills. Yet, rate control emerges as a single factor. It may be that there are day-to-day activities performed by many individuals which are similar to the tasks performed for rate control tests, and that "practice" on these rate control-like tasks explains how these different component abilities come to emerge as a unitary psychomotor factor.

Given the possibility that some of the abilities in Fleishman's psychomotor ability taxonomy may be comprised of several more specific abilities, one might wonder whether a more molecular approach to the psychomotor ability taxonomy would be most appropriate. For example, advocates of an information processing approach to psychomotor assessment argue that measurement of basic information processing abilities holds great promise for predicting performance. (See the summary of experimental and information processing research on psychomotor skills in the next section.) Implicit in the positions of many information processing researchers is the belief that the information processing conceptualization of human abilities is the most theoretically appropriate approach to ability assessment because the basic information processing abilities serve as the building blocks of all of the more complex abilities (Carroll, 1976). Indeed, many information processing researchers believe that some day they will be able to explain all complex abilities in terms of their component information processing abilities (cf. Sternberg, 1981).

Still, for several reasons, Fleishman's psychomotor ability taxonomy seems preferable to an information processing taxonomy for purposes of the current selection and classification research.

First, on a theoretical level, it would be naive to say that the abilities Fleishman has identified via factor analysis are somehow theoretically deficient simply because they represent a complex amalgamation of a number of more basic ability processes. Most major theories of cognitive abilities posit a hierarchy of abilities ranging from very basic processes such as encoding, decoding, and storage to a broad, general intellectual

ability factor (i.e., g). There is no reason to suspect that psychomotor abilities could not also be so structured. For example, in folk parlance, people are often described as "coordinated" and "handy" or "uncoordinated" and "klutzy." This broad psychomotor dimension seems analogous to g from the cognitive domain. Thus, the selection of an appropriate level of psychomotor ability complexity for the current research is really a pragmatic question: Which level of complexity is likely to result in measures with the highest predictive validity?

Second, assessment of information processing abilities has been plagued with psychometric problems. Allen, Rose, and Kramer (1978) have conducted research on the stability of individual differences in information processing abilities. Results of their research are summarized in Table 3. The reliabilities of these information processing measures tend to be rather low. Test-retest reliabilities (one-day interval between test sessions) ranged from .12 for Cluster Level on the Sentence Recall Task to .85 for both Inclusion Errors and Exclusion Errors on the Physical Match Task, with a median reliability of only .50. Other researchers have found that the stability of many information processing measures increases with practice (Sternberg, 1981). That is, the correlation between subjects' scores on Trials 1 and 2 of an information processing test would typically be quite a bit lower than the correlation between Trials 29 and 30. This lack of initial stability has led many information processing researchers to conclude that each subject should be allowed to practice an information processing task for several hours before assessment of the subject is attempted. In contrast, the reliabilities reported for the more complex psychomotor tests in Appendix A, which were obtained after allowing the subject little or no practice, are quite a bit higher. Almost all of the reliabilities are .70 or greater. Since the criterion-related validity of a test is limited by its reliability, the complex psychomotor tests are more likely to yield significant predictive validity coefficients if time constraints make it impractical to allow subjects a great deal of practice time prior to assessment.

Finally, and perhaps most importantly, in a validity study the appropriate level of ability complexity should be determined primarily by the complexity of the criterion one is attempting to predict (Hogan & Eagan, 1983). Listed below are some of the criteria from the current research which have been judged by experts to be dependent on psychomotor proficiency:

- o Repair Mechanical Systems. Perform corrective actions on previously diagnosed malfunctions of mechanical equipment or mechanical components using appropriate tools (e.g., wrenches, screwdrivers, gauges, hammers) in conjunction with technical information.
- o Construct Wooden Buildings and Other Structures. Perform carpentry activities (e.g., measure, saw, nail, plane) to frame, sheath and roof buildings, or to erect trestles, bridges, piers, and so forth.
- o Load Field Artillery or Tank Guns. Manipulate breech controls and handle ammunition (stow and load) to prepare guns for firing.

Table 3

One-Day Test-Retest Reliability of Information
Processing Abilities

<u>Information Processing Measure</u>	<u>Reliability</u>
Physical Match Task	
Inclusion Errors	.85
Exclusion Errors	.85
Letter Recall Task	
Order Recall	
Slope	.21
Intercept	.51
Derived Free Recall	
Slope	.19
Intercept	.49
Series Recall	
Slope	.24
Intercept	.41
Mental Addition Task	
Blanks	
Slope	.61
Intercept	.39
Correct	
Slope	.35
Intercept	.14
Sentence Recall Task	
Mean Cluster	.37
Maximum Cluster	.41
Cluster Level	.12
Sentence Recognition Task	
Errors	.18
New-Consistent	.14
New-Inconsistent	.27
Old-Consistent	.30
Old-Inconsistent	.31

(Continued)

Table 3 (Continued)

One-Day Test-Retest Reliability of Information Processing Abilities

<u>Information Processing Measure</u>	<u>Reliability</u>
Letter Rotation Task	
Inclusion Errors	
Slope	.53
Intercept	.72
Exclusion Errors	
Slope	.75
Intercept	.64
Set Membership Task	
Inclusion Errors	
Slope	.81
Intercept	.66
Exclusion Errors	
Slope	.79
Intercept	.66
Scan and Search Task	
Inclusion Errors	
Clear	
Slope	.76
Intercept	.77
Degraded	
Slope	.71
Intercept	.70
Exclusion Errors	
Clear	
Slope	.68
Intercept	.63
Degraded	
Slope	.60
Intercept	.55

Note. The data are from *An information processing approach to performance assessment: III. An elaboration and refinement of an information processing battery* (AIR-58500-TR) (pp. 56-57) by T. W. Allen, A. M. Rose, and L. J. Kramer, 1978, Washington, DC: American Institutes for Research. $N=54$.

- o Operate Track Vehicles. Use various vehicle controls to drive track vehicles (e.g., tanks, APCs, scout vehicles, bulldozers). Steer in response to terrain features.
- o Provide Medical and Dental Treatment. Give medical attention to soldiers in the field, or medical or dental clinic, or to animals (e.g., CPR, splinting fractures, administering injections, dressing wounds).

The description of these criteria suggests that performance on these tasks is determined in large measure by relatively complex psychomotor abilities. For example, the definition of control precision (ability to make fine muscular movements necessary to position equipment control mechanisms in response to a stimulus whose speed and/or direction of movement are perfectly predictable) bears great similarity to the description of some of the activities comprising the criterion Repair Mechanical Systems. Indeed, the history of many of the apparatus tests used by Fleishman (Melton, 1947) illustrates that these tests were designed specifically to tap abilities relevant to job performance in the military. Validity evidence, which will be presented in a subsequent section, indicates that these complex psychomotor tests have been quite effective in serving that function. Thus the complexity of Fleishman's ability factors appears perfectly appropriate given the objectives of the current research.

Construct Explication of the Psychomotor Domain

Anastasi (1983) has noted that the construct explication of a trait or ability requires input from several sources in addition to factor analysis. These sources might include field studies, naturalistic observation, task analyses, validity studies, controlled experiments, and psychophysiological or behavior genetic research.

Unfortunately, to date, there has been little effort expended on explication of the abilities comprising Fleishman's psychomotor taxonomy. Indeed, there have been few attempts to review and summarize our knowledge concerning any psychomotor ability or taxonomy.

In large measure, this lack of knowledge can be traced to decisions by the military in the mid and late 1950s to abandon psychomotor testing in the selection of pilots and flight crews (North & Griffin, 1977; Passey & McLaurin, 1966). Ironically, these decisions were made during the period that Fleishman was conducting his research into the taxonomy of psychomotor abilities.

Research conducted during World War II had proven the validity of psychomotor tests for predicting job and training performance for several military occupational specialties. There were, however, substantial problems with psychomotor test administration, especially for the apparatus tests. For example, most of the apparatus tests required individual administration, making them quite expensive to use and administer. Perhaps even more troublesome was the notorious unreliability of the various apparatus used in the administration of the tests. The Navy, for example, had decided during World War II that it would not use any selection tests which

could not be administered easily in decentralized testing sites (Viteles, 1942). Since subjects' test scores often varied significantly from one rudder control apparatus to the next, from one complex coordination apparatus to the next, etc., it was impossible for the Navy to ensure standardization of psychomotor apparatus tests among its test sites. Therefore, the Navy ceased to use apparatus tests. The Air Force followed suit 10 years later.

Initially, some psychologists believed that new, more reliable psychomotor tests could be developed to take the place of the original ACB apparatus tests. Adams (1956) attempted to develop a series of very simple motor tasks which he hoped would tap the same abilities as the ACB tests. Others worked on paper-and-pencil tests of psychomotor abilities. Virtually all of the validity results were disappointing. In a summary of these efforts, Cronbach (1970) concluded that it was unlikely that either paper-and-pencil tests or simple motor tests could measure the complex dexterity and coordination abilities that the apparatus tests measured.

Passey and McLaurin (1966) were among the first to advocate renewed use of psychomotor apparatus tests for pilot selection. They reviewed both the perceptual-psychomotor ability literature and the literature on the behavioral functions of a pilot. The goal of their review was to develop a list of ability constructs to guide future test development efforts for pilot selection. Their final list consisted of 19 ability constructs, including one called psychomotor ability. They felt that the most relevant psychomotor abilities for pilot performance would be those involving fine, highly controlled adjustment (i.e., control precision) and multilimb coordination. They also expressed interest in a construct they called anticipatory behavior. This construct was quite similar to Fleishman's rate control ability factor (Fleishman, 1958; Fleishman & Hempel, 1956). Passey and McLaurin recommended that future aircrew selection tests consist of "complex behavioral tasks involving intellectual, sensory-motor, and perceptual components" (p. 94).

Imhoff and Levine (1981) have also reviewed the perceptual, psychomotor, and cognitive ability literature as part of an effort to develop a list of ability constructs for inclusion in a pilot selection battery. Their review represents one of the most thorough efforts to explicate the ability constructs comprising the psychomotor domain.

Imhoff and Levine focused their summary of the psychomotor ability literature on two major bodies of research: the individual differences research conducted by Fleishman and his associates and research on the role of feedback in motor skills learning conducted by a number of motor skills learning researchers and theorists. Based on their readings, Imhoff and Levine concluded that most of Fleishman's psychomotor ability factors could be collapsed into two major dimensions.

The first dimension, basic movement speed and accuracy, subsumed Fleishman's control precision, speed of arm movement, and reaction time abilities. According to Imhoff and Levine, the movements controlled by these abilities tend to be repetitive and patterned. Once these movements are initiated there is little need for feedback or situational cues, since the movements require virtually no ongoing regulation (Laszlo, 1967; Laszlo & Manning, 1970). Little cognitive processing is required, and the

movements are typically quite short in duration. Keele (1968) concluded that these movements are probably under the control of motor programs. This led Imhoff and Levine to conclude that control precision, speed of arm movement, and reaction time were probably instrumental in both the development and the execution of these programs.

Imhoff and Levine called their second major psychomotor ability dimension perceptual-motor movement control. Included were multilimb coordination, response orientation, and rate control. According to Imhoff and Levine, these three abilities are important for "movement [which is] guided by sensory and perceptual feedback from initial responses" (p. 18). Movements controlled by this ability dimension tend to take longer to execute than the movements controlled by the basic movement speed and accuracy dimension. Moreover, each "movement" under perceptual-motor movement control actually consists of a number of short movements. The outcomes associated with each of these short movements may have important implications for the nature and timing of subsequent movements. Thus, cognitive processing of proprioceptive and environmental stimuli plays a key role in perceptual-motor movement control. For example, Fleishman and Rich (1963) found that subjects classified as high in kinesthetic sensitivity performed increasingly better over 10 trials on a two-hand coordination task relative to a group classified low on kinesthetic sensitivity. Other modes of feedback which have been shown to affect performance on tasks controlled by perceptual-motor movement abilities include visual (Lackner, 1974), auditory (Karlovich & Graham, 1968), and spatial orientation (Weitzman, 1979).

Imhoff and Levine's taxonomy represents an important contribution to the psychomotor ability literature. Their attempt to integrate motor skills learning theories and research with Fleishman's taxonomy represents one of the few attempts to explicate the psychomotor ability domain. More will be said about the relationship between these two important bodies of research in the next section.

While Imhoff and Levine's taxonomy provides useful information about the relationship among Fleishman's psychomotor ability factors, their ability dimensions may be too broad to be useful in describing jobs or in distinguishing between jobs or criterion constructs. As noted previously, the level of specificity in Fleishman's abilities corresponds closely to the level of specificity of the criterion constructs being used in the current research. Moreover, since Imhoff and Levine were concerned only with the psychomotor abilities relevant to pilot performance, their taxonomy failed to consider several important psychomotor abilities, including finger dexterity, manual dexterity, wrist-finger speed, aiming, and arm-hand steadiness. The use of these abilities as predictors of performance in the Army deserves consideration.

A Final Word on the Taxonomy of the Psychomotor Domain

Thus, in spite of the many criticisms which have been leveled against it and the alternatives which have been suggested to replace it, the psychomotor ability taxonomy identified by Fleishman (1975) would seem to be the one most relevant to the current Army research project. The level of ability specificity seems to correspond quite closely to the complexity of the major criterion tasks identified for this project. Moreover, the tests used in the identification of the taxonomy have been used successfully to

predict military job and training performance in the past (Melton, 1947), suggesting that the abilities underlying performance on these tests are indeed related to the criteria we are attempting to predict.

The one place where this taxonomy departs from Fleishman's (1975), as noted in the introduction to this section, is that reaction time and response orientation have been included in the cognitive-perceptual domain instead of in the psychomotor domain. Our research team will, of course, continue to evaluate the utility of these ability constructs as predictors for the current research. Thus, this departure from Fleishman's taxonomy should not be viewed as a rejection of the relevance and utility of these abilities. Rather, it should be regarded as a simple disagreement concerning the domain in which these abilities should be classified and discussed.

SECTION III

MOTOR SKILLS LEARNING

While the focus of this report is on individual differences in psychomotor abilities, the vast majority of psychological research on motor performance has been conducted by experimental psychologists interested in general laws and rules of motor learning. While very little of the research on motor skills learning has been directly concerned with the assessment or correlates of individual differences in psychomotor abilities, the motor skills learning literature still provides a great deal of information which is useful in understanding psychomotor abilities. For example, the various theories of motor skills learning which have been proposed have provided an improved understanding of the relationship between psychomotor abilities and cognitive, perceptual, and motor processes. Some of the research inspired by these models was cited in the preceding section in an effort to define and clarify the taxonomy of the psychomotor ability domain.

A Definition of Motor Skills Learning

Motor skills learning is "the acquisition of a temporal-spatial organization of muscular movement in a precise and consistent manner" (Hall, 1982, p. 248). The "temporal organization of muscular movement" refers to the fact that most motor movements require precisely timed muscular contractions and/or relaxations, while the "spatial organization of muscular movement" refers to the need to contract and relax the appropriate muscles to the precise degree required to achieve the outcome desired from the motor movement.

While motor skills learning is typically inferred from motor performance, learning and performance are not synonymous--not in motor skills nor in any other learning domain. That is, what a subject is observed to do is not always indicative of what a subject is capable of doing. Adams (1954), for example, required subjects to align a panel of lights via manipulation of a hand and foot control. The task was timed, with some subjects permitted more time to complete the alignment than others. Adams found that the subjects who were given the most time to complete the alignment made the most correct alignments per 2-minute trial, while the subjects who were given the least amount of time made the fewest correct alignments per trial. Subsequently, Adams removed all time limits and allowed subjects to complete the alignments at their own pace. He found no differences between groups in the number of alignments completed per trial, suggesting that the amount of learning in all groups was equal.

Typical Motor Performance Learning Curves

Figure 1 shows typical learning curves for four basic motor skills learning research paradigms. Figure 1A shows that, as subjects practice a task, they make fewer and fewer errors until performance reaches an asymptote near zero errors. Thus, performance becomes more accurate with practice. Figure 1B shows that the number of correct responses per trial increases with practice as performance becomes faster and/or more accurate (i.e., less time is expended on inaccurate responses). Research by Crossman (1959) indicates that number of correct responses per trial may

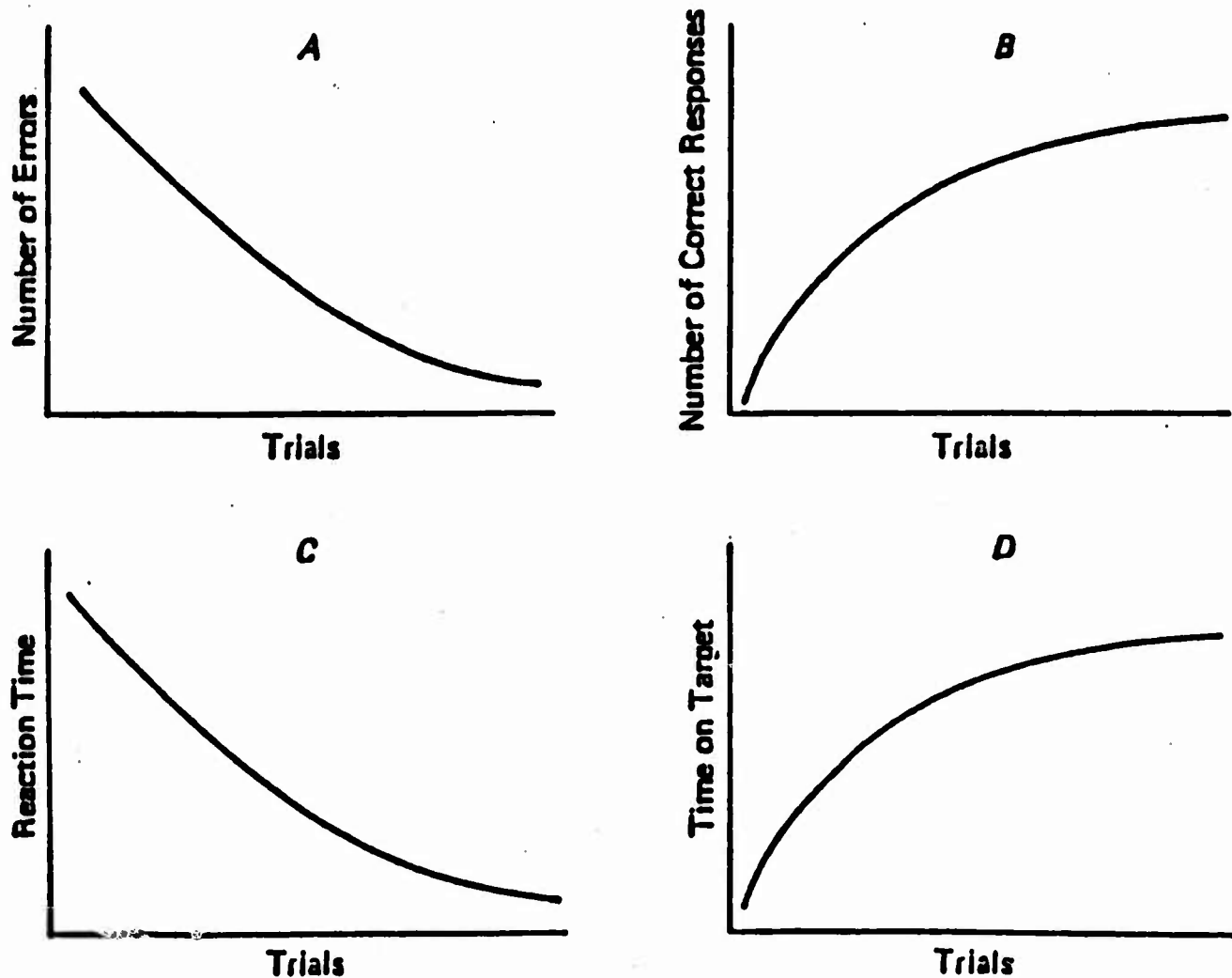


Figure 1. Typical learning curves for four basic motor skills learning research paradigms.

not reach an asymptotic level for years. Crossman found that workers in a cigar factory were still increasing their rate of cigar production after three years--or two million cigars--on the job. Figure 1C shows that subjects respond more quickly on a motor task with practice, until response time approaches zero. For example, Seibel (1964) found that his subjects' simple reaction time was still decreasing and showing no signs of reaching asymptote, even after 75,000 trials on a simple reaction time task. Figure 1D applies primarily to tracking tasks. A common dependent variable in these tasks is time on target. The figure illustrates that time on target increases (i.e., performance becomes more accurate) with practice. Indeed, all four figures illustrate how performance improves, but improves at a decreasing rate, as practice progresses until subjects' performance reaches a maximal, asymptotic level.

The Course of Motor Skills Learning

Stages of Motor Skills Acquisition. Bryan and Harter (1897) studied a group of subjects who were learning to use Morse Code. The researchers noted that subjects' performance improved consistently throughout many trials until their level of performance reached an apparent asymptote. Subsequently, following several trials where performance improved very little, performance again began to improve steadily. Bryan and Harter dubbed this the plateau phenomenon. They hypothesized that during the performance plateau, subjects were altering their strategies for performing the task. While subjects' initial strategies were efficient during the first trials of task learning, Bryan and Harter theorized, the strategies had to be altered in order for performance to become more automated and efficient.

Since Bryan and Harter's initial discovery, the plateau phenomenon has been replicated by a number of researchers. Some of these researchers have suggested that the phenomenon is artifactual. According to these researchers, the phenomenon appears to occur only because of the insensitivity of our measuring devices; learning actually occurs evenly throughout all skill acquisition trials. Most researchers, however, agree with Bryan and Harter's view that motor skills learning occurs in stages.

The most widely cited stage model of motor skills acquisition is Fitts and Posner's (1967). According to Fitts and Posner, learning occurs in three stages.

During the first stage, the subjects attempt to understand the nuances of the task and the desired outcome. Much of the subjects' efforts are invested in thinking and processing information. They must determine which situational cues are relevant to task execution. They must learn the motor steps involved in task performance. Often, they will say these steps aloud, as if for guidance as they perform the task. Because this phase of the task is distinguished by a considerable degree of cognitive activity, Fitts and Posner called it the cognitive stage.

During the second stage of task learning, the subjects' activities shift from "what to do" to "how to do it." The subjects attempt a number of different responses in an effort to find the most efficient means of achieving the desired outcome. Erroneous responses are abandoned, more successful response elements are combined with successful elements of other

response efforts, and the subject begins to develop an effective, integrated motor skill. This emphasis on experimentation and integration prompted Fitts and Posner to call this the associative stage of motor skills acquisition.

Entering the final phase of motor skills acquisition, the subjects are already quite proficient at the task. During this phase, performance will only improve somewhat. Nevertheless, important changes occur in the manner in which the subjects accomplish the task. The subjects learn to perform the task proficiently without paying much attention to what they are doing. This frees them to concentrate on any other tasks at hand. In addition, the subjects begin to use internal kinesthetic and proprioceptive cues as feedback to tell them whether they are performing the task correctly. The former reliance on external sources of feedback and information is diminished. To reflect the lessened attentional demands of the task during this stage, Fitts and Posner named this the automatic or autonomous stage of motor skills acquisition.

The Role of Abilities in Skills Acquisition. Stage models of motor skills acquisition are particularly interesting to differential psychologists because they suggest that the abilities contributing to skilled performance may vary depending on the skill acquisition stage. Fleishman and his colleagues have been among the most active researchers in pursuing this area of overlap between experimental and correlational psychology.

At the heart of Fleishman's research is the distinction he and his associates draw between abilities and skills. According to Parker and Fleishman (1960), "The term ability refers to a more general, stable trait of an individual which may facilitate performance in a variety of different tasks" (p. 1). For example, manual dexterity may be related to tasks as diverse as building a birdhouse and filling a cavity in a tooth. The term skill, on the other hand, refers to an individual's level of proficiency on a particular task. One might speak, for example, of an individual who is skilled at building birdhouses. This skill may be dependent upon a number of different abilities (e.g., manual dexterity, finger dexterity, numerical facility).

All theories of motor skills learning recognize that individuals become proficient at complex tasks via practice. Fleishman and other differential psychologists note, however, that even with extended practice individuals will differ greatly in their proficiency on a given task. Moreover, individuals will also differ in their proficiency on a task during their very first effort at performing that task. In an effort to understand this phenomenon more clearly, Fleishman and his colleagues have investigated a number of features of the skills acquisition process. The research questions they have addressed include:

1. What abilities are related to performance on complex tasks during the initial phases of task practice and learning?
2. What abilities are related to performance on complex tasks during the final phases of task practice and learning (i.e., when individuals have reached an asymptotic level of task proficiency)?

3. How stable are individual differences in task proficiency from the initial to the final phase of task practice and learning?
4. Do the abilities contributing to individual differences in task proficiency during the initial phase of task practice and learning differ from those contributing to individual differences during the final phase of task practice and learning?
5. In comparison with other parameters governing skilled task performance (e.g., massed vs. distributed practice, amount of time elapsed since the task was last practiced), how important are individual ability differences in determining an individual's level of task proficiency?

To answer these questions, Fleishman and his associates conducted several studies (e.g., Fleishman, 1960; Fleishman & Fruchter, 1960; Fleishman & Hempel, 1954b, 1955; Parker & Fleishman, 1960). In a typical study, these researchers administered a battery of psychomotor and cognitive-perceptual reference tests to a sample of 200-300 subjects. Subsequently, subjects received extended practice on a complex criterion task. Factor analysis was used to identify the abilities represented in the reference test battery. The factor loadings of the various criterion task practice trials on these ability factors represented the correlations between the abilities and the criterion task at various stages of practice.

Fleishman (1967, 1972, 1975) has reported four general conclusions based on this research:

1. As practice continues, the particular combination of abilities contributing to task performance changes.
2. These ability changes are progressive and systematic from trial to trial. Eventually the changes cease and the relative contributions of the various abilities to task performance stabilize.
3. Psychomotor abilities become relatively more important determinants of task performance (vis-a-vis cognitive-perceptual abilities such as spatial or verbal ability) as practice continues.
4. As practice continues, an ever-increasing percentage of task performance variance is specific to the task itself (i.e., it is unrelated to any of the broad, general abilities represented in the reference test battery).

These conclusions have been bolstered by other studies conducted by Fleishman and his colleagues using different research designs. In one of these studies, Fleishman (1957) performed two separate factor analyses for a test battery. For one factor analysis, the test battery data included scores from a complex task taken from the initial stages of practice on that task. For the second factor analysis, the test battery included

scores for the same task from a later stage of practice. In two other studies, Fleishman (1965) and Fleishman and Fruchter (1965) examined the relationship between total task performance and performance on task components at various stages of practice. Data from all of these studies were consistent with the four basic conclusions noted above.

Other researchers have also investigated the relationships between psychomotor skills and abilities. Kohfeld (1966) found that correlations between performance on a complex psychomotor task and a verbal test decreased consistently over 15 practice trials on the task, while correlations between the task and a psychomotor test increased with task practice. Adams (1953, 1957) and Hinrichs (1970) also found that the ability correlates of a complex psychomotor task changed as the task was practiced. Unlike Fleishman, however, Hinrichs found that task performance was equally predictable from the ability tests during early and late practice trials. According to Hinrichs, there was no evidence that task-specific variance increased as the task was practiced. Hinrichs attributed this finding to the fact that he included ability tests which closely resembled the complex task in his test battery. Hinrichs found that the correlations between these tests and the task actually increased as the task was practiced. This finding is certainly not inconsistent with Fleishman's finding that variance specific to a task increases with practice.

Models of Skill-Ability Relationships. Three different models have been developed in an effort to explain Fleishman's four major research findings summarized previously. These models were described and evaluated in a 1971 technical report by Hulin and Alvares.

The changing task model, which Fleishman (1966) espoused, suggests that the abilities required for proficient task performance change with practice. An individual's level of task proficiency on a given practice trial thus depends on both the abilities possessed by the individual and the degree to which each of these abilities contribute to task performance at that stage of practice.

Adams (1957), Humphreys (1960), and others have offered a competing model, the changing subject model. While there are important differences in the changing subject models offered by the various advocates of this position, all of these models suggest that practice on a complex task actually changes an individual's ability level. It is therefore no surprise that correlations between the task and ability test scores taken before the task was first practiced should decrease as practice continues; the practice is actually changing the individual's true ability level. One of the most important differences between the changing task and changing subject models is that the changing subject model makes no distinction between skills and abilities. According to the changing subject model, changes in task proficiency are changes in ability.

The third model represents a combination of the first two models (Hulin & Alvares, 1971). Like the changing subjects model, this third model makes no distinction between skills and abilities. Thus, changes in task proficiency are attributed at least in part to an individual's improving abilities. The third model also postulates, however, that the abilities involved in successful task performance change as practice on the task continues.

According to Hulin and Alvares, data on psychomotor abilities and psychomotor task proficiency provide no clear indication of which model is most accurate. None of the three models can be rejected based on currently available data.

The validity of these three models is important because the models have differing implications for training. The changing task model, for example, suggests that the most appropriate training strategy is to focus training on the abilities involved in task performance during each stage of practice. In a test of this hypothesis, Parker and Fleishman (1961) used two different training strategies in teaching subjects to perform a complex tracking task. One of the strategies was a "common sense" strategy designed to resemble typical military training and instruction. Subjects assigned to this condition received a verbal description and demonstration of the task. They were then allowed to practice the task individually. During this practice, they received guidance and assistance from instructors. The training curriculum for the "experimental" group was designed to provide subjects with an opportunity to improve abilities which were known to be related to task performance at various stages of practice. For example, previous analysis of the task revealed that spatial orientation was most highly related to task performance during the fourth tracking session. Therefore, subjects were provided with special instructions designed to acquaint them with the spatial orientation requirements of the task prior to and during the third tracking session. An analysis of tracking error scores revealed that the experimental group learned the tracking task more quickly and attained a higher level of tracking proficiency. These differences in tracking performance persisted even after the two groups began to receive identical training beginning with the 11th tracking session.

Additional Research on the Role of Individual Differences in Learning.

Fleishman and his colleagues have also investigated the relationship between individual difference variables and several other learning and training phenomena.

Fleishman and Rich (1963) showed that, as practice continues, correlations between performance on a two-hand coordination task and spatial-visual abilities decreased while correlations between task performance and kinesthetic abilities increased. These findings suggest that the spatial cues which subjects use to learn how to perform a task diminish in importance as the task is learned. Instead, subjects begin to rely more on proprioceptive feedback (i.e., knowledge of the "right feel" or the "right touch") as they become more proficient.

In a study on skill retention, Fleishman and Parker (1962) examined the reliability of individual differences in task proficiency. Initially, Fleishman and Parker administered a battery of ability tests to their subjects. Subjects then practiced a complex psychomotor task extensively for seven weeks before a period of no practice was instituted. This no-practice period lasted either 1, 4, 9, 14, or 24 months. At the conclusion of the no practice period, subjects again completed the task. Fleishman and Parker found that task performance during the final set of practice trials correlated very highly with performance on the task after the long no-practice period (r 's ranged from .80 to .90). None of the ability

measures contributed additional significant variance to the prediction of of task performance following the no-practice period. Fleishman and Parker therefore concluded that subjects' level of retention was based solely on the habits they had acquired during practice.

A final example of Fleishman's research on the impact of individual differences on learning phenomena is a study by Fleishman and Ellison (1969) on interference/transfer and massed vs. distributed practice effects. (The next part of this section contains a somewhat more detailed discussion of transfer and massed vs. distributed practice effects.) Subjects in this study completed a battery of spatial, psychomotor, and personality tests. They then received both massed and distributed practice trials on a complex psychomotor task. Following extended practice, the controls on the task apparatus were shifted. Subjects completed several trials on this altered apparatus. Finally, the apparatus was reconfigured to its original state and subjects completed one last trial.

Results of the study showed that a few of the ability measures were somewhat useful in predicting which subjects would experience transfer or interference effects when the apparatus was altered. None of the correlations was very great, however. In addition, there was no evidence that any of the personality measures were useful in predicting which subjects would show "flexibility" and transfer. One of the most interesting findings from this study concerned the prediction of performance on the first trial *following* a massed practice trial. On virtually every trial, multiple R squared between task performance and the ability measures was between .40 and .50. During the trial immediately following massed practice, though, multiple R squared dropped to .20. In their evaluation of this result, Fleishman and Ellison suggested that the ability requirements for task performance during massed and distributed practice trials are identical. Ability requirements appear to change following massed practice, however, perhaps because of the "fatiguelike" state induced by massed practice.

Variables Affecting Motor Skills Learning

Cognitive, perceptual, and psychomotor abilities represent individual differences variables affecting the rate of acquisition of and ultimate level of proficiency in skilled motor performance. Most researchers, however, have focused on how situational and motivational variables affect motor skills learning across individuals. Much of this research has centered on five key variables: massed vs. distributed practice, part vs. whole task learning, stimulus-response compatibility, knowledge of results, and feedback.

Massed vs. Distributed Practice. Massed practice refers to a condition where subjects practice a single task repeatedly with no intervening periods of rest. The opposite is distributed practice, where practice trials are interrupted by rest periods or periods of practice on other tasks.

In a classic study, Ammons (1950) provided rest intervals of zero, 20, or 50 seconds; 2, 5, or 12 minutes; or 24 hours between thirty-six 20-second practice trials on a rotary pursuit task. Ammons found that the time on target scores of the massed (i.e., zero-second rest interval) practice group were considerably lower than those of the distributed

practice groups throughout all 36 trials. There was no apparent relationship between the length of the rest period and time on target scores, however, suggesting to Ammons that all forms of distributed practice are equally superior to massed practice in facilitating motor skills learning. This finding has been replicated in numerous studies.

More recent research has focused on the persistence of massed vs. distributed practice effects. Whitley (1970) administered 25 trials of a tracking task to two groups of subjects. One group received massed practice, the other distributed practice. From the fifth trial through the last, the performance of the distributed practice group grew increasingly superior to the performance of the massed practice group. This finding replicated the results of much previous research. Whitley continued his experiment, however. Following the 25th trial, he gave both groups a 5-minute rest period. Then he administered 10 more trials to each group. During these 10 trials, there was no difference in the performance of the two groups. Moreover, both groups performed at a skill level comparable to the pre-rest skill level of the distributed practice group. Thus, the performance of the massed practice group showed a considerable improvement from the final pre-rest trial to the first post-rest trial.

Based on these results, Singer (1980) stated that "there is little reason to doubt the superiority of distributed practice over massed practice for immediate performance in a variety of tasks" (p. 420). He added, however, that there is no evidence that *skill retention* is facilitated by distributed practice. The common belief that distributed practice is superior to massed practice is not clearly supported by currently available data.

Part vs. Whole Task Learning. Many complex tasks can be taught by breaking the task into subtasks or parts and then teaching the task one part at a time. This is known as part task learning or the part method of learning. This contrasts with whole task learning, in which the subject attempts to learn to perform the task as an integrated whole. Another variant is the progressive-part or continuous-part procedure, in which the subject practices the first subtask first, then the first two subtasks together, then the first three subtasks together, and so forth.

Psychologists and educators have long been interested in determining which type of task learning is most conducive to skill acquisition. Results from this body of research have been mixed, however, with no clear evidence in favor of either part or whole task learning (Hall, 1982; Singer, 1980).

In an effort to explain these apparently contradictory results, Naylor and Briggs (1963) tested the hypothesis that the appropriate method of learning depends upon the complexity and organization of the task. Task complexity refers to the information processing demands of the task. Complex tasks require subjects to think a lot about what they are doing, to remember information from previous experiences or practice trials, to make judgments about appropriate methods for performing the task given continuously changing environmental cues, and so forth. Task organization refers to the interrelationships among the various dimensions or subtasks comprising the task. Highly organized tasks consist of highly interrelated and interdependent subtasks. A task with low organization consists of

several relatively independent subtasks. The Naylor-Briggs model suggests that part task learning is superior when tasks are high in complexity and low in organization, while whole task learning is superior for tasks which are low in complexity and high in organization.

While the model has some intuitive appeal, there are no data to support or refute it at this time. One problem with evaluating the model is that it is often difficult to rate the organization and complexity of a task. For example, organization and complexity ratings for the task "swinging a golf club" are highly unreliable (Singer, 1980). In addition, even when experts do agree that a task is best taught in parts, they often differ considerably on how to divide the task into parts for part task learning. This again makes comparison of part and whole task learning difficult. Thus, the current state of affairs in this area provides almost no general rules for determining when and where part task learning will be superior or inferior to whole task learning.

Stimulus-Response Compatibility. One variable with a tremendous impact on motor skills learning is the compatibility between the stimulus and the response. Stimulus-response compatibilities are derived from stereotyped behavioral response patterns shared by nearly everyone within a population. For example, the turn signal in an automobile operates in a compatible manner (Hall, 1982). To signal a turn to the right, one must push the turn signal clockwise, to the right. Conversely, one must push the turn signal counterclockwise, to the left, to signal a left turn. This task is learned quickly by almost everyone who is learning to drive a car. It is likely, however, that reversing the operation of the signal (i.e., so that a left turn would be signaled by pushing right on the turn signal) would result in an incompatible stimulus and response and therefore greatly increase the difficulty of task learning.

The example above illustrates the potential benefits of designing tasks so that stimuli and responses are compatible. Part of the AAF Aviation Psychology Program during World War II was devoted to such research (Wickert, 1947). Alarmed by the high loss of lives and planes caused by pilots' spatial disorientation during combat missions, the AAF directed the Aviation Psychology Program to investigate possible solutions. Psychologists used the critical incident technique to gather data on disorientation experiences from pilots (Flanagan, 1954). These data helped psychologists identify a number of stimulus-response incompatibilities within the cockpit of an airplane. Psychologists involved in this study were subsequently able to issue a number of recommendations which led to improvements in the design of cockpits, controls, and instrument panels in combat airplanes.

Two apparatus have been used extensively in laboratory investigations of stimulus-response compatibility.

The first is a choice reaction time apparatus (see Figure 2). Subjects begin each trial by resting their index finger on the home button. One of the four stimulus lamps is then lit. Normally, a subject responds to this stimulus by moving the index finger and depressing the button adjacent to the lighted lamp. When the machine is in an incompatible mode, though, the assignment of stimulus lamps to response buttons is random. That is, the response button assigned to a particular stimulus lamp is not necessarily the button adjacent to that lamp.

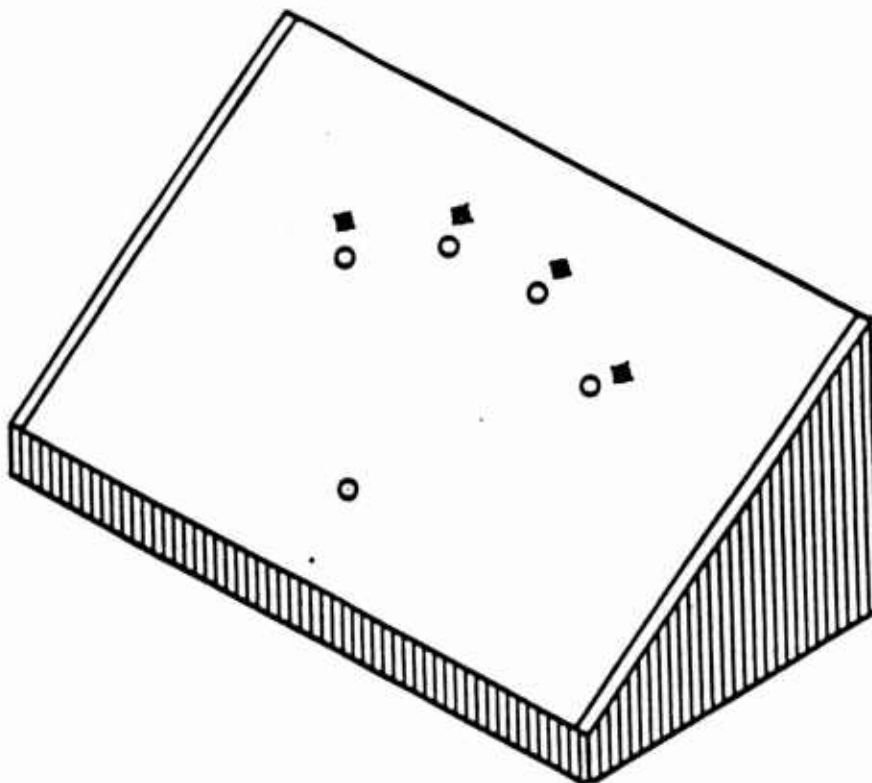


Figure 2. Four-choice reaction time apparatus.

The second apparatus (Hall, 1982) consists of eight pathways or spokes emanating from a central hub (see Figure 3). Adjacent pathways are separated by 45° angles. At the end of each pathway are lamps that can be used as test stimuli. A digital clock-like device mounted atop the apparatus can also be used to present stimuli. At the beginning of each trial, the subject must place a metal stylus in the center of the hub. When the apparatus is in compatible mode, the subject must trace with the stylus through the pathway in the direction corresponding to the stimulus presented. For example, if the lamp at the end of the right horizontal pathway was lighted (or if "3:00" was presented on the clock-like device), the subject would trace with his stylus across the right horizontal pathway as quickly as possible. When the stimulus is in an incompatible mode, however, assignment of responses to stimuli is random. Thus, a display of "3:00" might correspond to a response of tracing down the lower vertical pathway.

Not surprisingly, results from both the reaction time and the pathway apparatus show that subjects' reaction times are much greater when the stimulus and response are incompatible. The results from the second apparatus are interesting, however, because they illustrate how the strength of the compatibility between a stimulus and a response affects reaction time when the compatibility is eliminated. In one study involving this apparatus, Fitts and Deininger (1954) found mean reaction times of .35 second and .64 second for the lamps and clock, respectively, when the apparatus was in compatible mode. These data suggest that the compatibility was greater for the lamps than for the clock. When the machine was in incompatible mode, however, reaction times were faster for the clock than for the lamps (.77 second vs. .96 second). Thus, the data show that interfering with a highly

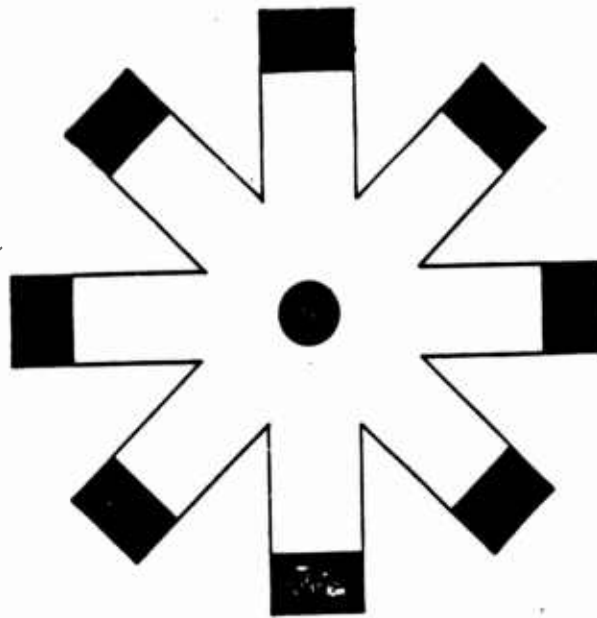


Figure 3. Pathway apparatus testing reaction time.

compatible stimulus-response relationship causes large decrements in skilled motor performance. The resulting incompatible responses are even more difficult to learn than the incompatible responses resulting from tasks which, in a compatible mode, had relatively weaker stimulus-response relationships.

Knowledge of Results. Knowledge of results (KR) refers to information about the effectiveness of task performance received from external sources. In some cases, a subject receives KR by observing or otherwise sensing (e.g., smelling, listening to) the *results* of his actions. In other instances, the subject obtains KR from other individuals (e.g., an experimenter).

Thorndike (1927) demonstrated the power of KR in a very simple experiment. He blindfolded subjects and asked them to draw lines which were either three, four, five, or six inches long. He then provided some subjects with KR. That is, when the subject's line was within one-eighth inch of the desired length, Thorndike would say "right." Otherwise, Thorndike would say "wrong." Thorndike provided no KR to a control group. Results indicated that the group receiving KR became significantly more accurate in their drawing during the course of the 600 trials, while the control group showed no improvement.

In subsequent studies, researchers have varied a number of KR parameters to determine their effects on motor skills learning. Results for some of the most important parameters are summarized below.

Several studies have shown that the frequency of KR is related to task performance. In one study (Bilodeau & Bilodeau, 1958), subjects were given KR either every trial, every third trial, every fourth trial, or every tenth trial. Bilodeau and Bilodeau found that increased frequency of KR was related to reductions in the degree of task performance error.

Increasing the amount and precision of KR also promotes motor skills learning. During a tracking task, Smode (1958) provided one group of subjects with continuous KR while a second group was told about their performance only after each tracking trial had been completed. Mean time on target was significantly greater for the first group. In a similar experiment, Hardesty and Bevan (1964) recorded the performance of four groups on a reaction time task. One group received no feedback (control), a second group was told only that their performance was "very good," "too slow," and so forth after each trial (qualitative), a third group was told their reaction time in seconds after each trial (quantitative), and a fourth group received both qualitative and quantitative KR after each trial. The qualitative-quantitative group had the fastest mean reaction time, followed by the quantitative group, the qualitative group, and finally the control group.

The effects of a delay in KR have been studied extensively. Results show that the nature and timing of the delay determine how the delay will affect motor skills learning and task performance. Simply delaying KR for a period of several seconds has no effect on performance (Lorge & Thorndike, 1935). If, however, KR is delayed so that one or more trials intervene between the target trial and KR, then task performance is adversely affected. Moreover, the number of trials intervening between the target trial and KR is positively related to the degree of decrement in task performance (Bilodeau, 1956). The amount of time elapsing between KR and the onset of the next trial also affects task performance. Weinberg, Guy, and Tupper (1964) investigated the effects of post-KR delay intervals of 1, 5, 10, and 20 seconds. They found that performance was poorest in the group which received only a one-second KR delay. They found no performance differences among the three remaining groups. Weinberg et al. concluded that subjects require more than one second to process KR effectively. With only a one-second KR delay, subjects are unable to use KR to improve their performance.

A final parameter of KR which has received considerable attention is withdrawal of KR. Newell (1974) asked his subjects to move a slide 9.4 inches along a track in 150 milliseconds. At the completion of each trial, Newell asked his subjects to estimate their movement time. Newell used six KR groups. One group received KR on every trial. For the other five groups, KR was withdrawn after 2, 7, 17, 32, or 52 trials. For all six groups, whenever KR was provided, it was provided only after subjects had estimated their movement time. Subjects in all six groups completed 75 trials of the movement time task. In general, Newell found that movement time tended to depart from 150 milliseconds after KR was withdrawn. The discrepancy was smaller, however, for subjects who received extensive KR before KR withdrawal. Moreover, the group which received KR for 52 trials before KR withdrawal showed no performance decrement following withdrawal. Newell concluded that extensive KR creates internal standards which allow subjects to judge the accuracy of their performance without input from external sources. His analysis of subjects' estimated movement times

confirmed this. Subjects who received extensive KR gave very accurate estimates of their movement times, even after KR withdrawal, while subjects who received little KR gave very inaccurate estimates of their movement times.

Feedback. When motor skills researchers speak of feedback, they refer to a variable that is distinct from KR. KR denotes information arising from external sources. KR is based on outcomes occurring outside the subject. Feedback, on the other hand, arises from sources within the subject. Feedback is based on proprioceptive stimulation. For example, subjects may sense how far they have moved an arm, how hard they have squeezed a peg, or how much they have bent a wrist. These sensations are all examples of proprioceptive stimulation.

In the next part of this section, some of the major theories of motor skills learning will be presented. The role of feedback in motor skills learning is a major source of disagreement among these theories. Therefore, a discussion of research results from feedback studies will be postponed and integrated into the presentation of the competing theories below.

Theories of Motor Skills Learning

Several theories of motor skills learning have been offered to explain the course and stages of motor skills acquisition. These theories consist primarily of descriptions of the processes occurring within an individual during motor skills learning. Thus, the focus of these theories has been on motivational variables (e.g., KR and feedback) rather than on task or stimulus variables (e.g., massed vs. distributed practice, part vs. whole task learning, stimulus-response compatibility).

Habit Theory. Perhaps the oldest and simplest theory of motor skills learning is the theory that learning builds a habit state in an individual. When the habit state is sufficiently strong, habit theory suggests that the behavior will occur reliably.

The weakest part of habit theory is that it fails to consider many interesting learning phenomena. For example, habit theory states that most motor skills learning occurs through KR. Yet, KR alone is insufficient to explain why motor skills learning occurs in stages. KR provides no information about the processes underlying these stages. Nor is habit very useful in explaining transfer. Transfer refers to the fact that motor skills learning that has occurred in one situation can facilitate motor skills learning in related situations (Ammons, Ammons, & Morgan, 1958). Habit suggests no mechanisms for such a phenomenon.

In sum, habit theory is not so much wrong as deficient theoretically. As a psychological construct, habit fails to provide insight into the mechanisms and processes underlying motor skills learning. Thus, motor skills researchers have moved beyond habit theory in their efforts to describe and explain motor skills learning.

Cybernetic Theories. Cybernetics is the study of control systems in machines and animals. It is concerned with how machines and humans regulate themselves in order to maintain internal systems within standards of performance (i.e., by ensuring that errors do not exceed limits of

tolerance). Perhaps the most familiar example of a cybernetic control system is the bimetallic thermostat. The metal in the thermostat acts as a switch to turn a furnace on and off. Heat and cold cause the metal to expand and contract, respectively, regulating whether the furnace is off or on. No intervention from outside sources is required to control the furnace; the thermostat is constructed to be self-regulating.

The key to any cybernetic control system is feedback (i.e., information arising from within the machine or organism). For this reason, cybernetic control systems are often called closed-loop systems. With respect to motor skills, this means performance is evaluated and regulated primarily on the basis of internal information, such as proprioceptive and kinesthetic feedback. The cybernetic control theory of motor skills learning which has received the most attention is Adams' (1971) closed-loop theory.

According to Adams, during the early stages of motor skills learning, the subject conceptualizes motor learning as a problem to be solved. He uses KR to guide him in solving this problem. He tries new responses, receives and processes KR, and adjusts his responses accordingly. As the subject practices the task, he receives feedback from internal sources as well. This feedback might include proprioceptive stimuli from receptors in joints and muscles, tactual stimuli from nerve cells in the fingers, and kinesthetic stimuli based on body position and movement. These stimuli form a reference mechanism which Adams calls a perceptual trace. As performance becomes more proficient, the subject's perceptual trace becomes stronger and a more valuable source of information. During any given attempt to execute a motor response, a subject can compare feedback from ongoing proprioceptive stimulation to performance standards embedded within the perceptual trace. This allows the subject to regulate his performance more efficiently than if he were forced to wait and attend to only KR. Thus, in the closed-loop system, feedback gradually replaces KR as the key regulator of skilled motor performance.

Besides perceptual trace, the other major construct in Adams' closed-loop theory is memory trace. The memory trace contains the information needed to initiate a movement. It is activated by appropriate environmental cues, and then initiates and specifies the speed and direction of movement. The memory trace is not a rigid motor program, however. Once the movement is initiated, the perceptual trace and feedback provide the subject with the means to regulate and alter any movement initiated by the memory trace.

Adams' theory is very useful in explaining the transition from the associative to the automatic stage of motor skills learning. The theory can also account for the finding that withdrawal of KR has no effect on performance if the subject has received extensive KR prior to KR withdrawal (Newell, 1974). According to Adams, the reason performance does not suffer following KR withdrawal is that the subject's behavior has become self-regulating.

There is some research evidence, though, which suggests that skilled motor performance can occur in the absence of any proprioceptive feedback. In a classic experiment involving laboratory rats, Lashley and Ball (1929) severed the nerves carrying the proprioceptive stimuli from muscles and

joints to the motor centers of the brain after the rats had learned to run a maze. Following the operation, the rats' motor coordination was somewhat impaired. Nevertheless, the rats' ability to traverse the maze correctly was almost totally unaffected. Skilled performance persisted even in the absence of feedback. Lashley (1951) has also pointed out that a pianist's fingers can strike up to 16 keys per second accurately, which appears to be much faster than a human closed-loop control system could operate (cf. Adams, 1976). Thus, many researchers lean towards a theory of motor skills learning which posits more centralized control of movement.

Hierarchical Control Theories. Adams' closed-loop theory stresses the role of feedback in providing ongoing regulation of movement. Other theorists have questioned whether humans actually regulate their movements to the extent Adams has suggested. Lashley's research (Lashley, 1951; Lashley & Ball, 1929) provides at least two examples where closed-loop control of movement is either impossible or highly implausible. Lennenberg's (1967) finding that speech may require up to several hundred muscle movements per second provides another example of motor performance which would seem to occur too rapidly to be much affected by feedback.

In an effort to account for these findings, several researchers have offered open-loop theories of motor skills learning. The theories are open-loop in the sense that they provide no mechanism for ongoing regulation of movement. The theories all suggest that once movement is initiated by central processes it is rarely, if ever, affected by feedback. Thus, according to these theories movement control and regulation is part of the domain of higher-order processing functions.

Keele (1968) has proposed a motor program theory of motor skills learning. Keele would concur with Adams that KR and feedback produce both perceptual and memory traces. In Keele's theory, however, the memory trace is more important than the perceptual trace in motor movement. The memory trace is not simply a mechanism for initiating movement. Instead, the memory trace represents a fairly detailed specification of the muscle movements required to perform a particular task. Each motor program runs for 200 to 400 milliseconds (Schmidt, 1975). Once a program is initiated, it is virtually impossible to alter movement until the program runs its course.

Keele (1973) indicates that the chief role of feedback in motor skills learning is in the development of motor programs. Feedback, along with KR, helps to create and perfect motor programs. Once a program is functioning efficiently, feedback does not affect the functioning of that program during motor movement. If new information (i.e., from KR) suggests that the program requires alteration, feedback will aid in modifying the program. Otherwise, the role of feedback in motor skills learning is ended once the motor program is functioning properly.

Keele's theory is particularly applicable to brief movements lasting for but a fraction of a second. These movements are often called ballistic acts. Keele (1968) has theorized that complex movements can also be understood in terms of his motor program theory, however. According to Keele, complex movements are nothing more than a series of ballistic acts. During the course of such movements, ongoing visual, tactile, auditory, and kinaesthetic stimulation help determine the timing and sequence of ballistic

acts. In this manner, both feedback and KR play a crucial role in complex movement. Yet, at the most basic level, the ballistic acts themselves are under the exclusive control of motor programs and cannot be altered during the course of movement execution.

Evidence in favor of the motor program theory comes primarily from two sources. Evidence from research with human subjects cited above (e.g., Lennenberg, 1967) suggests that some movement occurs too rapidly to be under continuous, ongoing regulation. Adams (1976), however, has presented evidence from physiological studies which indicates that information can be processed through feedback loops in as little as 10 milliseconds. Thus, this first source of evidence may not be as powerful as once thought. The second source of evidence comes from animal studies (e.g., Delcomyn, 1980; Lashley & Ball, 1929; Wilson, 1961). Results from these studies suggest that elimination of proprioceptive stimulation is not accompanied by degradation of performance. Closed-loop theorists have questioned the applicability of these results for humans, however.

Adams, Goetz, and Marshall (1972) have conducted one of the only tests of the competing predictions of closed-loop and open-loop theories. During the course of motor skills learning, Adams et al. changed the conditions of feedback. For one group, the feedback change occurred early in practice. For a second group, the change occurred late in practice. According to Adams et al., closed-loop theory predicts that feedback change would have its greatest impact late in practice, when performance is being closely regulated by feedback. Open-loop theory predicts that feedback change would have little effect on performance late in practice because movement during this stage is controlled primarily by motor programs. Adams et al. found that performance degradation was much greater when the feedback occurred late in practice, providing support for Adams' closed-loop theory.

Information Processing Theories. The most common application of information processing theories has been to provide models for the investigation of basic cognitive processes such as attention, perception, encoding, storage, retrieval, and decision making. As the name implies, information processing research concerns the manner in which individuals attend to and process sensory stimuli.

Most information processing research has been concerned with determining the limits of human cognitive processing capabilities. Craik and Lockhart (1972), for example, investigated limitations in the ability to store information in short-term storage. Broadbent (1958) described limits in the ability to attend to two or more stimuli simultaneously.

Figure 4 presents an information processing model devised by Welford (1968). Welford's model is particularly interesting for two reasons.

First, Welford's model illustrates clearly the measurement dilemma faced by information processing researchers. In a typical information processing experiment, researchers present an "external object" to the subject. Subsequently, they measure some overt response produced by an "effector." These overt responses are the only measurable output available. From these responses, information processing researchers attempt to infer the nature of the cognitive processes occurring within the subject. Welford's model in Figure 4 represents a rather simple information

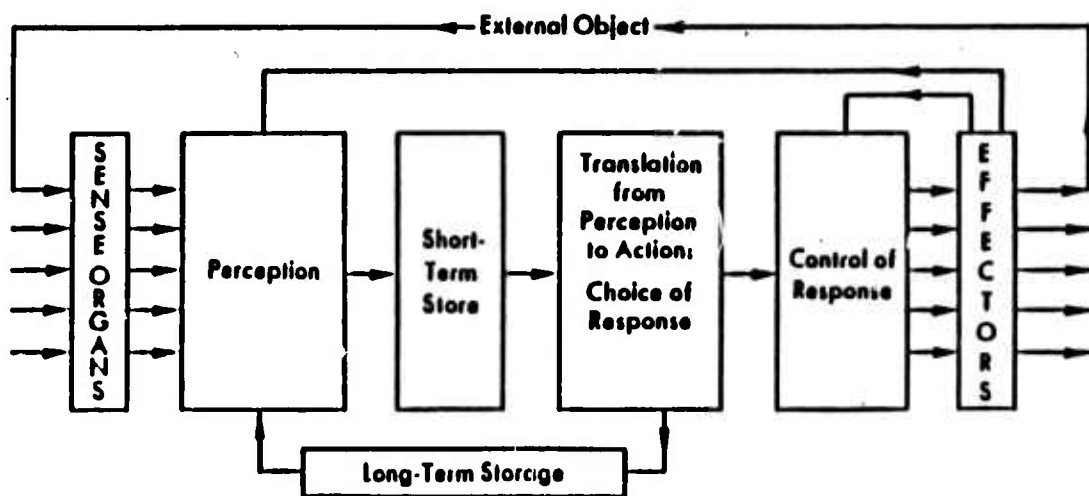


Figure 4. An information processing model of motor skills learning. From *Fundamentals of skill* by A. T. Welford, 1968, London: Methuen. Copyright 1968 by Methuen. Reprinted by permission.

processing model--and it includes five processes intervening between the stimulus and the response. None of these processes are measurable. Thus, experimental tests of information processing hypotheses often must rely on measurement of phenomena occurring several steps removed from the target process. This frequently makes it difficult to test information processing hypotheses adequately.

Second, Welford's theory represents one of the few attempts to link motor skills learning to information processing theory. The processes Welford calls "choice of response" and "control of response" are especially relevant to skilled performance.

Welford's information processing theory is primarily descriptive, especially with respect to the processes involved in selecting and executing motor responses. Thus, it is difficult to compare his model directly with Adams' or Keele's. Welford has been particularly interested in investigating how attentional limitations affect motor response time. For example, he has described how responses to a signal are delayed if the signal is presented immediately following another signal (Welford, 1976). Based on this and other research, Welford has concluded that central processing mechanisms must be heavily involved in motor responses. When these mechanisms are attending to other stimuli, as in the example above, response time to motor movement cues is lengthened considerably.

An Integrative Model of Motor Skills Learning. Singer (1980) has noted that cybernetic, hierarchical control, and information processing theories tend to focus on different aspects of the motor skills learning process and thus are more complementary than incompatible. He has developed an integrative model of motor skills learning that borrows elements from all three theories.

Information processing theory is primarily concerned with the cognitive processing that occurs during skilled performance. Welford's (1968, 1976) work has demonstrated how attentional limits affect motor response time and performance. During the cognitive stage of motor skills acquisition, for example, subjects must pay attention to the task at hand in order to detect key features of the task that might be helpful in determining an appropriate motor response strategy. Also during this stage, attention is vital to the storage of task information in long-term memory. Transferring task information from short term memory to long term memory can only occur as the subject rehearses (i.e., attends to) the task information (Craig & Lockhart, 1972). Even after motor skills learning reaches the automatic stage, attention continues to play a key role in motor performance. As Welford (1976) showed, if the subject fails to attend to situational information or is in the process of responding to other cues and signals when new situational information is presented, motor responses either will not occur or will be delayed significantly.

Adams' (1971) cybernetic closed-loop theory is useful in describing the increasing importance of feedback and the perceptual trace as motor skills acquisition passes through the associative and automatic stages. Adams' theory also explains the important role KR plays in the learning process. Research by Adams et al. (1972) demonstrates that, at least under certain conditions, feedback continues to function as a regulator of task performance even after a task has been mastered by the subject.

Keele's (1968) hierarchical control theory explains most motor functioning during the automatic stage of motor skills acquisition. During this stage, there appears to be very little regulation of performance; instead, most motor movement appears to be pre-programmed. Information processing is still important during this learning stage, however, because the subject must determine the proper sequence and timing to execute these motor programs.

Singer's (1980) integrative model suggests a number of fruitful areas for future research. Much more research is needed to describe the role of attention and decision making during the cognitive stage of motor skills acquisition. A description of the cognitive processes involved in sequencing and timing the ballistic acts comprising complex movements is another area ripe for research. Of course, the roles of KR and feedback during the automatic phase of motor skills acquisition continue to be hotly debated. Singer's model indicates that these also deserve consideration in future research.

Implications of Motor Skills Learning Theories for Psychomotor Ability Research

In the preceding section, it was noted that a major problem with Fleishman's ability taxonomy is that the abilities have never been described and explicated thoroughly. In developing definitions of these abilities, Fleishman attempted to judge the common ability demands of the tests loading on each ability factor. By 1962, Fleishman had conducted sufficient factor analytic research to allow him to specify a comprehensive psychomotor ability taxonomy. Since that time, however, very little research has been directed at developing a nomological net for these abilities.

Recognizing this problem, Imhoff and Levine (1981) showed how findings from research on motor skills acquisition could be used to identify and explain the relationships among Fleishman's abilities. For example, they noted that the type of motor performance described by Keele's (1968) open-loop theory was similar to the types of tasks performed in measures of control precision, reaction time, and speed of arm movement. This implies that two of the most important processes involved in these abilities are detection (i.e., to determine when to initiate a movement) and movement speed (i.e., to permit rapid responding). Keele's description also suggests some processes that would be relatively unimportant to these abilities. These include monitoring feedback and KR, since the movements involved in measures of these abilities are under the control of motor programs.

Previously, it was noted that Imhoff and Levine built their psychomotor taxonomy around the role of feedback in task performance. Singer's integrated model suggests a number of other processes and parameters that might be considered when describing Fleishman's abilities. These include:

1. The role of situational information in initiating movement.

2. The difficulty of detecting situational information (e.g., due to weak or competing stimuli).
3. The attentional demands of task learning.
4. The attentional demands of task performance when motor skills acquisition has reached the automatic stage.
5. Sources of KR and feedback.

Imhoff and Levine's review reflects their expert judgment of the relationship between motor movement processes and parameters and Fleishman's psychomotor abilities. Empirical research is needed to supplement and validate these judgments. Such research could be modeled after the cognitive correlates, cognitive components, cognitive training, and cognitive contents research paradigms described by Sternberg (1981). Eventually, such research might indicate improved ways of categorizing psychomotor abilities. Certainly it would improve our understanding of how these abilities operate.

SECTION IV

THE USE OF PSYCHOMOTOR TESTS IN APPLIED PSYCHOLOGICAL RESEARCH

For both legal and pragmatic reasons, organizations using selection tests must be concerned with the validity of those tests. The most widely accepted method of establishing the validity of a selection test involves demonstrating that test scores are related to one or more indices of job performance. This is known as criterion-related or empirical validity. According to Kleiman and Faley (1978), since the passage of the 1964 Civil Rights Act, the courts have tended to rely almost totally on criterion-related validity evidence in determining the validity of a selection test or procedure. In response to the courts' position, industrial psychologists have also focused their validation efforts on establishing the criterion-related validity of their selection tests. The criterion-related validity evidence for psychomotor tests will be evaluated in a subsequent part of this section.

Kleiman and Faley, however, argued that the current emphasis on criterion-related validity is erroneous. Because of small sample sizes and inadequate personnel research staffs and budgets, many organizations lack the resources to conduct quality criterion-related validity research. In such instances--and even in instances where criterion-related validity research is economically feasible--Kleiman and Faley suggested that the validity of selection tests can be established via content validation.

Content Validation of Psychomotor Tests

Content validity is established by demonstrating that a test constitutes a representative sample of the behaviors comprising the domain to be measured (Anastasi, 1976; Helmstadter, 1964). With respect to selection testing, this means that one must demonstrate that the content of a test is a representative sample of the behaviors comprising the job performance domain. Generally, the content validity of a selection test is established by conducting a thorough job analysis, extracting situations from the job performance domain for use as test items or stimuli, assessing subjects' responses to those situations, and evaluating those responses against known standards of job performance.

Many psychologists have noted that content validity alone does not ensure that a test will be valid (i.e., that a test will do what it is intended to do) (Cascio, 1982; Guion, 1977). This is particularly applicable for selection tests, whose primary or sole function is to predict how well job applicants will perform on the job.

Nevertheless, the principles of and procedures for content validation can be useful in guiding test development. Guion (1977), for example, argued that content validity helps ensure that test items reflect an appropriate operational definition of the constructs or behaviors comprising the criterion domain. Wernimont and Campbell (1968) distinguished between tests as signs and tests as samples of behavior. In their view, tests consisting of job samples and work samples are more directly related to the criterion of interest (i.e., they are more content valid) than tests that are signs or indirect measures of the criterion. Because of this direct relationship, Wernimont and Campbell predicted that the criterion-related

validity of "samples" tests should generally be greater than that of "signs" tests.

In sum, content validity alone does not ensure that a test will be useful for predicting job performance. It does seem reasonable, however, to expect that, on the average, content valid tests will have greater criterion-related validity than tests which lack content validity.

Anastasi (1976) has stated that psychomotor tests are content valid if they require the subject to "closely reproduce all or part of the movements required in the performance of the job itself....(T)o ensure validity, the test and the job should call for the use of the same muscle groups" (p. 444).

The identification of the movements and motor skills required for successful job performance is dependent on careful job analysis. As noted above, job analysis should always be the first step in the content validation process for selection tests (Equal Employment Opportunity Coordinating Council, 1978).

The AAF aircrew selection research that was described briefly in the taxonomy section represents the first systematic effort to apply the principles of content validation to the development of psychomotor tests. While their job analysis (Miller, 1947) was crude by current standards, it did serve to focus the psychomotor test development efforts of Melton (1947) and his colleagues.

Pilots were the target of the first AAF job analysis. An early variant on Flanagan's (1954) critical incident technique was used to collect job analysis data. Because the goal of this study was to predict attrition from pilot training, job analysis data collection focused on pilot training failures.

AAF procedures during World War II required check-pilots (i.e., pilot training instructors) to appear before a faculty board to justify their decisions regarding the elimination of cadets from pilot training (Melton, 1947; Miller, 1947). At these proceedings, the check-pilot informed the student and the board members of the reasons for recommending that the cadet be withheld from further training. Usually the check-pilot made reference to characteristics of the cadet and generated specific examples of problems the cadet was experiencing in learning how to fly. The check-pilot's statements and the board's comments and recommendations were recorded in the *Faculty Board Proceedings*.

In 1941, researchers began to analyze the *Faculty Board Proceedings* to determine common reasons for pilot training failure. Initially, proceedings were reviewed for 300 cadets eliminated from elementary pilot training. In 1942, a second analysis was conducted based on 1,000 additional cadets who had also been eliminated from elementary pilot training. Later, the analysis was expanded to include 100 cadets who had been eliminated from advanced single engine pilot training and 100 cadets who had been eliminated from advanced twin engine pilot training. Based largely on the analysis of the sample of 1,000 cadets who had been eliminated from elementary flight training, a list of 20 common reasons for pilot training failure was compiled. These 20 reasons were grouped into four major

categories: intelligence and judgment, alertness and observation, coordination and technique, and personality and temperament.

Of particular interest to those involved in psychomotor test development was the third category, coordination and technique. This category subsumed five reasons for pilot training failure: coordination, appropriateness of controls used, feel of the controls, smoothness of control movement, and progress in developing technique. Table 4 provides more complete explanations of these five reasons for pilot training failure. Table 4 also shows the percentage of cadets whose elimination was based at least in part on each of these reasons. Of the four major categories, coordination and technique was the most commonly cited failure category for cadets in elementary pilot training and the second most commonly cited failure category for advanced single and twin engine pilot training. Therefore, the AAF research team decided to emphasize the development of tests which would reliably assess individual differences in coordination and technique.

From the beginning, the AAF research team felt that it would be necessary to use apparatus tests to measure coordination effectively. Therefore, they invested much of their time in designing new apparatus tests which were similar in appearance and psychomotor demands to apparatus and instruments in the cockpit of an airplane. Among the tests developed by this team during World War II were the Two-Hand Coordination Test, the Two-Hand Pursuit Test, the Rudder Control Test, and the Arm-Hand Steadiness Test. In addition, these researchers modified and improved a number of tests which had previously been used in pilot selection research, including the Complex Coordination Test and the Rotary Pursuit Test. These tests all are described in Appendix A.

By maintaining fidelity between the tests and the job performance domain, the AAF researchers guaranteed the content validity of their psychomotor tests. Their efforts to validate these tests did not cease with content validity, however. They also evaluated the criterion-related validity of each test before including it in the psychomotor portion of the ACB. Moreover, the research team was careful to include new psychomotor tests in the ACB only if they tapped variance in pilot training attrition which had not been tapped by other psychomotor tests. This helped ensure coverage of a broad, representative sample of the criterion domain. Validity analyses for the ACB indicated that the psychomotor apparatus tests were tapping unique variance in the prediction of pilot training attrition --variance that was not tapped by traditional paper-and-pencil cognitive and perceptual tests.

Another example of content validation in psychomotor test development is provided by recent research into the prediction of tank crew performance (Campbell & Black, 1982; Eaton, Johnson, & Black, 1980). Based on their analyses of tank crew jobs, these researchers identified several tasks that are crucial to tank crew performance. They then developed a series of work sample tests using these key tasks as guides. Thus, their tests represent content valid "samples" rather than "signs" of criterion behavior. Like the AAF research team, these researchers followed up their test development efforts by collecting criterion-related validity data. Analyses of these data indicated that several of the psychomotor work sample tests correlated

Table 4

Percentage of Pilot Training Failures Attributable in Whole or in Part to Problems Involving Coordination and Technique

Reason for Failure	Type of Training		
	Elementary (N=1,000)	Single Engine (N=100)	Twin Engine (N=100)
COORDINATION AND TECHNIQUE (overall) Includes any reference to at least one of the five reasons for failure listed below	81%	91%	69%
Coordination Inability to apply the correct pressures to the controls either in combination or in the proper sequence with proper timing	58%	74%	37%
Appropriateness of Controls Used Lack of knowledge of control or controls to use, and inability to operate those controls so as to achieve the desired attitude of the plane	21%	10%	13%
Feel of the Controls Inability to sense the responsiveness of the plane to control movement, or the effect being produced on the plane by various control pressures; inability to detect the "stiffness" or "mushiness" of controls	2%	23%	3%
Smoothness of Control Movements Inability to operate controls with smooth, even movements and good touch control, without evidence of roughness, heaviness, or jerkiness	22%	30%	25%
Progress in Developing Technique Slowness in learning the various coordinations of the controls and techniques necessary for flying the plane	54%	52%	34%

Note. From Apparatus Tests (Army Air Forces Aviation Psychology Program Research Report No. 4) (p. 62) by A.M. Melton (Ed.), 1947, Washington, DC: U. S. Government Printing Office.

significantly with training performance measures, especially for the tank gunner job.

These two studies, along with many others (e.g., Eggenberger, 1976; Fowler, 1981; Jones, 1982), demonstrate how measures developed using a content validation strategy can be useful in predicting training performance and job proficiency. It must be recognized, however, that content validity alone does not ensure that a selection test will be a valid predictor of job performance. Many selection researchers, including Cascio (1982) and Guion (1977), have advised that content validity is best viewed as a test development strategy rather than as a test validation strategy (cf. Kleiman & Faley, 1978). In fact, Guion (1978) has even suggested that a more appropriate name for content validity would be content-oriented test development.

Thus, under no circumstances should content validity be regarded as a substitute for criterion-related validity. No matter how careful researchers are in designing job selection tests, there is no guarantee that these tests will be valid as predictors of job performance criteria. Therefore, criterion-related validity data should be collected and evaluated whenever possible. Indeed, the ultimate goal of any validation study should be collection of criterion job performance measures and calculation of the correlation between these measures and selection test scores.

As an interim step before criterion-related validity can be estimated, researchers can be guided in their choice of selection tests by previous criterion-related validity research. Recent investigations of validity generalization by Dunnette et al. (1982) and Pearlman, Schmidt, and Hunter (1980) suggest that, for any given job family, correlations between selection tests and job performance measures vary only slightly from job to job and from organization to organization. As an aid in determining which selection tests are likely to be valid predictors of performance for particular jobs, a summary of previous validity research is provided next.

The Criterion-Related Validity Literature Review Process

Recently, Bullock and Svytanek (1985) showed that the results of a meta-analysis or validity generalization study can be greatly influenced by the manner in which a researcher chooses to record the characteristics and results of studies under review. Therefore, a thorough description of the methods used to identify the literature on the criterion-related validity of psychomotor tests and to summarize the criterion-related validity coefficients reported for the psychomotor domain is provided below.

Identifying Relevant Articles, Reports, and Manuals. In the Preface to this report, procedures used to identify the literature relevant to this domain were described. Initially, the review of criterion-related validity studies focused on research conducted within the past 15 years (i.e., prior to 1970). As the sources identified via the computerized literature search were reviewed, however, it became apparent that very little criterion-related validity research had been conducted in the psychomotor area during that period. Indeed, with the exception of several recent military research studies investigating the validity of psychomotor tests as predictors of pilot performance (e.g., Hunter & Thompson, 1978; Imhoff & Levine, 1981; Myers, Jennings, Schemmer, and Fleishman, 1982) and some

ongoing validity research with the General Aptitude Test Battery (GATB) (U. S. Department of Labor, 1970), there have been very, very few research programs devoted to an investigation of the criterion-related validity of psychomotor tests. Moreover, of the military research programs listed above, those initiated by Imhoff and Levine for the Air Force Human Resources Laboratory (AFHRL) and Myers et al. for the Army Research Institute (ARI) have yet to yield any validity results. Thus, excluding the GATB, it was only possible to accumulate approximately 300 validity coefficients from the studies initially reviewed.

Consequently, the literature search was expanded to include validity research published during the 1940s, 1950s, and 1960s. This made it possible to incorporate several additional major sources of psychomotor validity data into the review, including pre-1970 GATB validity results and validity data from the AAF selection research on pilots, navigators, and bombardiers from World War II (Melton, 1947).

In summarizing all of the articles, reports, and manuals reviewed for the psychomotor domain, the research staff completed 180 article review forms and 422 predictor review forms. Not all of the papers reviewed contained criterion-related validity information, however; only 35 articles, reports, and manuals reported criterion-related validity coefficients. (References for these 35 studies are provided in Appendix B.) In total, these 35 studies yielded 2,373 validity coefficients based on test scores from 75 different psychomotor apparatus and paper-and-pencil tests. By far the largest set of these coefficients ($N=1,734$, or 73.1%) were obtained from the five GATB psychomotor scales. An additional 350 coefficients (14.7%) were obtained from the Army Air Forces aircrew selection research conducted during World War II (Melton, 1947). These 350 coefficients represent research results for 32 different psychomotor tests.

Tabulation of Validity Coefficients. During the review, it was noted that different authors tended to report validity results differently. For example, some researchers who investigated several different samples of subjects reported validity coefficients for each sample while others simply reported a mean coefficient across samples. Some attempted to correlate psychomotor test scores with rather global criteria (e.g., supervisory ratings of overall job performance) while others reported separate validity coefficients for each of several more specific criteria (e.g., performance on several different work sample tests). Finally, for tests involving multiple trials, some researchers reported separate validity coefficients for each trial while others computed a total score across trials and based their validity coefficient(s) on this total score.

To ensure consistency in the tabulation of validity coefficients, the following decision rules were adopted:

1. If validity coefficients were reported for more than one sample of subjects, separate validity coefficients were recorded for each sample.
2. If validity coefficients were provided for more than one criterion variable:

- a. If there was a total or summary score, only the validity coefficient(s) for that summary score was recorded.
- b. If there was no total or summary score, separate validity coefficients were recorded for each criterion variable.
3. If separate validity coefficients were recorded for each trial (or for each block of trials) on the psychomotor test:
 - a. If there was a total or summary score, only the validity coefficient(s) for that summary score was recorded.
 - b. If there was no total or summary score, the *mean* validity coefficient(s) across trials was computed and recorded.

Special forms were then prepared to facilitate the tabulation of validity data. The following information was recorded for each validity assessment:

1. Predictor construct
2. Criterion construct
3. Validity coefficient
4. Sample size
5. Research setting
6. Job type
7. Pilots vs. non-pilots
8. Test
9. Reviewer
10. Article review form number
11. Predictor review form number

The predictor construct refers to the psychomotor ability tapped by the predictor test used in the validity assessment. Each test was assigned to one of 10 different predictor constructs. Nine of the constructs were taken from Fleishman's taxonomy of the psychomotor ability domain: multi-limb coordination, control precision, rate control, finger dexterity, manual dexterity, wrist-finger speed, aiming, arm-hand steadiness, and speed of arm movement. As noted in the section reviewing the taxonomy of the psychomotor domain, except for reaction time and response orientation this list includes all of the abilities Fleishman included in his taxonomy of the psychomotor domain. Table 2 in the taxonomy section includes a definition of each of these nine abilities. The tenth predictor construct, complex psychomotor predictors, includes all tests which significantly tap two or more of the nine abilities listed above. The tests categorized into this predictor construct often involved complex tasks such as video games

(e.g., Jones, Kennedy, & Bittner, 1981) or job simulations (e.g., Eaton, Johnson, & Black, 1980).

The criterion construct refers to the type of criterion measure that was used to validate the predictor construct. Criteria were initially categorized into one of four major categories: educational and school achievement, training performance, job proficiency, and job involvement/withdrawal. Each of these major categories was in turn divided into 2-6 criterion constructs, resulting in 12 different criterion constructs. Definitions and explanations of these constructs are provided in Table 5.

For research setting, each study was categorized as either military or non-military. For the psychomotor domain, this distinction was based entirely on whether the subject population was military or civilian.

The job type classification scheme was derived from the *Dictionary of Occupational Titles (DOT)* (U. S. Department of Labor, 1977) and Ghiselli's General Occupational Classification system (Ghiselli, 1966). Ten broad job groups were initially identified from a review of these two sources. These job groups were then altered in light of information concerning important distinctions between different military occupational groups. For example, because mechanical maintenance and electronics comprise two broad, distinct job types within the Army, the DOT Structural occupation category was divided into these two job types. In total, nine job types were used to classify the psychomotor validity assessments. These job types are listed in Table 6. Table 6 also includes a short list of sample military and non-military jobs within each job type.

Since so many of the subjects used in psychomotor validity research were pilots or pilot trainees ($N=416$ validity coefficients, or 17.5% of all tabulated validities), each validity assessment was also classified as either pilot or non-pilot. This classification represents a finer breakdown of the professional job type from Table 6.

The test refers to the psychomotor test used in the validity assessment. As noted previously, 75 different psychomotor tests were used in the criterion-related validity studies reviewed. These tests are listed in Table 7. Table 7 also lists the psychomotor ability tapped by each test and provides a reference to an article in which the test has been used and/or described. Several of the more widely used tests listed in Table 7 are pictured and described in Appendix A. Appendix A also contains summary reliability and validity information for these tests.

The reviewer, article review form number, and predictor review form number were tabulated in order to identify the article, report, or manual from which each validity assessment was obtained.

Method of Summarizing Validity Information. Numerous tables summarizing the validity evidence for each predictor construct-criterion construct pair were prepared and are presented below. Within each predictor-criterion cell, several pieces of information are reported. The median validity coefficient appears as the first entry in the cell. This is followed by the weighted mean validity coefficient, which is the mean validity across all coefficients tabulated for that cell. In computing this mean, the validity coefficients were weighted by sample size. For

Table 5

Criterion Constructs

Major Category	Criterion Construct	Definition or Explanation
Educational and School Achievement	Grades	Academic course grades or GPA
	Instructor evaluations	Instructor ratings or rankings
Training Performance	Objective measures	Paper-and-pencil exam scores, achievement test scores, or course grades based solely on paper-and-pencil exams
	Subjective measures	Instructor ratings or rankings
	Combination objective and subjective measures	Final course grades based on paper-and-pencil test scores and instructor evaluations (Note: Unless it was specifically stated that training course grades were based on objective exams or subjective evaluations, they were categorized into this "combination" construct)
	Go-no go training courses	Pass/fail, graduate/non-graduate, or successful/unsuccessful outcomes or number of washbacks
	Hands-on measures	Work sample or job sample measures which are scored objectively or based on instructor evaluations

(Continued)

Table 5 (Continued)

Criterion Constructs

Major Category	Criterion Construct	Definition or Explanation
Job Proficiency	Subjective measures	Supervisor or peer ratings or rankings
	Job-related measures	Job knowledge or work sample tests
	Archival measures	Units produced, salary rates or increases, promotions, etc.
Job Involvement/ Withdrawal	Job satisfaction	Job satisfaction or attitude survey ratings
	Job withdrawal	Absenteeism, re-enlistment, or voluntary turnover

Table 6

Job Types and Sample Jobs

Job Type	Sample Military Jobs	Sample Non-Military Jobs
Professional, Technical, and Managerial	Air Force officers Pilots Navigators Intelligence	Managers, supervisors, foremen Engineers Health care professionals (e.g., dental hygienist) Pilots Draftsmen
Clerical	Office clerk Administrator Personnel specialist Communications specialist	Secretary Office clerk Switchboard/ keyboard operator Telegrapher
Sales	None	Sales representative Sales clerk
Protective Services	Military police Combat soldier Infantryman General enlisted personnel Undifferentiated apprentices	Police trainees Security guard Correction officer
Service	Food service Medical specialist	Food service Medical, dental assistant Truck driver
Mechanical and Structural Maintenance	Aircraft mechanic Vehicle mechanic Munitions mechanic	Machinist mechanic Carpenter Plumber Welder Appliance repairman

(Continued)

Table 6 (Continued)

Job Types and Sample Jobs

Job Type	Sample Military Jobs	Sample Non-Military Jobs
Electronics	Electronic and radio repairman Radar repairman Sonar technician Surveillance specialist Radio operator	Electronics repairman Electrical technology trainees
Industrial	None	Machine operator Processor, assembler, bench worker Ironworker Coal miner General maintenance worker
Miscellaneous	Attack submarine trainee	Power plant operator

Table 7

Psychomotor Tests Used in Criterion-Related Validity Research

Test Name	Psychomotor Ability Construct	Number of Articles, Reports or Manuals Reporting Validity Data	Number of Validity Coefficients Obtained	Reference ^b
Aetna Drivotron	Complex	1	3	Farr et al., 1971
Aircraft Landing Test	Complex	1	8	Fowler, 1981
Allstate Good Driver Trainer	Complex	1	3	Farr et al., 1971
Arm-Hand Steadiness Test ^a	Arm-Hand Steadiness	1	13	Melton, 1947
Arm-Leg Coordinator	Multilimb Coordination	1	1	Vant, 1962
Artificial Horizon Pursuit Test	Rate Control	1	2	Melton, 1947
Automated Pilot Aptitude Measurement System (APAMS)	Complex	1	22	Hunter and Thompson, 1978
Bennett Hand-Tool Dexterity Test	Manual Dexterity Multilimb Coordination	2	3	Rim, 1962
Bi-Manual Coordination Test	Multilimb Coordination	1	8	Melton, 1947
Career Determining Exercises	Complex	1	3	Farr et al., 1971
Center-of-Mass Task	Complex	1	8	Eaton et al., 1980
Chalk Carving Test	Complex	1	2	Mathews and Jensen, 1977
Compensatory Balance Test	Multilimb Coordination	2	5	Payne et al., 1952
Complex Coordination Test ^a	Multilimb Coordination	2	19	McGrevey and Valentine, 1974
Complex Coordination Test ^a	Multilimb Coordination	9	47	Melton, 1947
Conflicting Manipulation Test	Manual Dexterity Multilimb Coordination	1	2	Melton, 1947
Crawford Small Parts Dexterity Test	Finger Dexterity Aiming	1	14	de Wet, 1959
Crossing Test	Manual Dexterity	2	3	Grant and Gray, 1970
Custom Form Board	Manual Dexterity	1	1	Croll et al., 1973
Dial Setting Test	Control Precision Aiming	1	8	Inskeep, 1971
Dotting Test	Control Precision Aiming	1	2	Melton, 1947
			2	Melton, 1947

(Continued)

Table 7 (Continued)

Psychomotor Tests Used in Criterion-Related Validity Research

Test Name	Psychomotor Ability Construct	Number of Articles, Reports or Manuals Reporting Validity Data	Number of Validity Coefficients Obtained	Reference ^b
Electronic Pilot Pursuit Test	Multilimb Coordination	1	1	Melton, 1947
Formboard Test	Manual Dexterity	1	3	Farr et al., 1971
GATB Aiming Scale	Wrist-Finger Speed	1	26	U. S. Department of Labor, 1970
GATB Finger Dexterity Scale ^a	Finger Dexterity	2	561	U. S. Department of Labor, 1970
GATB Manual Dexterity Scale ^a	Manual Dexterity	2	561	U. S. Department of Labor, 1970
GATB Motor Coordination Scale ^a	Wrist-Finger Speed	2	554	U. S. Department of Labor, 1970
GATB Motor Speed Scale	Wrist-Finger Speed	1	26	U. S. Department of Labor, 1970
Gunner Tracking Task	Complex	1	2	Campbell and Black, 1982
Hand-Foot Reaction Test	Multilimb Coordination	1	4	de Wet, 1960a
Hayes Pegboard Test	Finger Dexterity	1	2	Shanthamani, 1978
Jump Reaction Time Test	Complex	1	3	Farr et al., 1971
Large Tapping Test ^a	Wrist-Finger Speed	2	2	Croll et al., 1973
Line Control Test	Arm-Hand Steadiness	1	1	Croll et al., 1973
Mark Making Test	Wrist-Finger Speed	1	2	Mathews and Jensen, 1977
Mashburn Serial Apparatus Test	Multilimb Coordination	1	9	Mashburn, 1934
Minnesota Rate of Manipulation Test ^a	Manual Dexterity	2	3	Kim, 1962
Motor Judgment Test ^a	Rate Control	1	6	Farr et al., 1971

(Continued)

Table 7 (Continued)

Psychomotor Tests Used in Criterion-Related Validity Research

Test Name	Psychomotor Ability Construct	Number of Articles, Reports or Manuals Reporting Validity Data		Number of Validity Coefficients Obtained	Reference ^b
		Validity Data			
Observer Trainability Test	Complex	1	1	1	Jones, 1982
O'Connor Tweezer and Finger Dexterity Test ^a	Finger Dexterity	4	12	12	Rim, 1962
Pedestal Sight Manipulation Test	Rate Control	1	4	4	Melton, 1947
Peg-Moving Test	Finger Dexterity	1	1	1	Melton, 1947
Peg Placing and Turning Test	Finger Dexterity	1	4	4	Mathews and Jensen, 1977
Pennsylvania Bi-Manual Work Sample	Manual Dexterity	2	9	9	Rim, 1962
Photoelectric Aiming Test	Arm-Mand Steadiness	1	1	1	Melton, 1947
Pinboard Test	Finger Dexterity	1	3	3	Farr et al., 1971
Pistol Firing Test	Complex	1	1	1	Osborn and Ford, 1976
Plane Control Test	Multilimb Coordination	1	2	2	Payne et al., 1952
Purdue Pegboard Test	Finger Dexterity	1	1	1	Rim, 1962
Pursuit Confusion Test	Control Precision	1	2	2	Payne et al., 1952
Pursuitmeter	Multilimb Coordination	1	1	1	Melton, 1947
Rotary Pursuit Test ^a	Control Precision	4	23	23	Melton, 1947
Round Adjustment Task	Complex	1	4	4	Eaton et al., 1980
Rudder Control Test ^a	Multilimb Coordination	8	19	19	Melton, 1947
Rudder Timing Reaction Test	Complex	2	9	9	Melton, 1947
Santa Ana Finger Dexterity Test ^a	Finger Dexterity	5	34	34	Melton, 1947
Single-Dimension Pursuitmeter ^a	Rate Control	1	4	4	Melton, 1947
Small Tapping Test ^a	Aiming	1	1	1	Croll et al., 1973
Steadiness Aiming Test ^a	Arm-Mand Steadiness	1	1	1	Melton, 1947
Steadiness Test	Arm-Mand Steadiness	1	5	5	de Vet, 1960b
Stromberg Dexterity Test	Manual Dexterity	1	1	1	Rim, 1962

(Continued)

Table 7 (Continued)

Psychomotor Tests Used in Criterion-Related Validity Research

Test Name	Psychomotor Ability Construct	Number of Articles, Reports or Manuals Reporting Validity Data		Number of Validity Coefficients Obtained	Reference ^b
		Validity Data			
Timing Reaction Test	Complex	2	19	Melton, 1947	
Trace Tapping I ^a	Aiming	1	1	Croll et al., 1973	
Trace Tapping II ^a	Aiming	2	2	Croll et al., 1973	
Tracing Test	Arm-Hand Steadiness	1	1	Croll et al., 1973	
Tracking Task	Complex	1	24	Melton, 1947	
Triform Pegboard Test	Finger Dexterity	1	5	Melton, 1947	
Two-Hand Coordination Test	Multilimb Coordination	1	6	Carstedt; cited in Melton, 1947	
Two-Hand Coordination Test ^a	Multilimb Coordination	2	8	McGrevey and Valentine, 1974	
Two-Hand Coordination Test ^a	Multilimb Coordination	6	46	Melton, 1947	
Two-Hand Pursuit Test ^a	Multilimb Coordination	1	8	Melton, 1947	
Two-Plate Tapping Test ^a	Speed of Arm Movement	1	1	Melton, 1947	
Variable Coordination Test	Multilimb Coordination	1	2	de Vet, 1962	

a. These 21 tests are described in more detail in Appendix A.

b. These references simply represent one article in which the test has been used and/or described. In many instances, tests have been used and described in more than one study.

example, a validity coefficient obtained from a sample of 100 subjects was weighted twice as heavily as a validity coefficient obtained from a sample of 50 subjects. The third entry is the number of different articles, reports, and manuals from which validity coefficients were obtained for that cell (K). To a certain extent, K represents the number of independent studies which contributed validity coefficients to that cell. It should be recalled, however, that the majority of validity coefficients summarized in the tables were taken from just two sources--the manual for the GATB (U. S. Department of Labor, 1970) and the AAF World War II research on aircrew selection (Melton, 1947)--and that these sources both reported validity results from many semi-independent research efforts. The fourth entry L tells the number of validity coefficients included in that cell, while the fifth entry M indicates the number of different psychomotor tests used in the validity assessments for that cell. Finally, the N range tells the range of sample sizes for the L validity assessments. The number in parentheses following this range is the median sample size across all of the studies included in that cell.

Initially, summary tables were prepared for each job type for both military and nonmilitary subject populations. Since there were no validity results for five military job types, a total of 13 tables were prepared. All of these 13 tables are presented and discussed below.

In addition, there were some instances where it was useful to combine different subject populations (e.g., across research settings or across job types) or to examine special subject populations (e.g., pilots) in more detail. Thus, there are many special summary tables interspersed among the 13 main summary tables below.

Finally, at the end of the validity summary section there is a table which summarizes the criterion-related validity evidence by predictor construct and by major criterion category (e.g., job proficiency) for each job type. This table shows only the median validity coefficient and the number of validity coefficients (L) for each cell.

The Criterion-Related Validity Evidence for Psychomotor Tests

Professional, Technical, and Managerial Jobs. Table 8 summarizes the criterion-related validity coefficients for military professional, technical, and managerial jobs. The validity coefficients in this table have been taken almost exclusively from military research on aircrew selection. The vast majority of these coefficients are for the job of pilot, while almost all of the remaining coefficients are derived from research on navigators.

Table 8 indicates that psychomotor abilities are quite powerful predictors of training performance for these jobs. Multilimb coordination and control precision appear to be particularly valid predictors. Median validities for multilimb coordination are in the .20s for two different training performance constructs, while the median validity of control precision for predicting graduation from training school is .22 ($N=20$ validity coefficients). In addition, several other psychomotor abilities show moderate validity for predicting graduation from training. The median validity for rate control is .12 ($N=6$ validity coefficients), the median validity for manual dexterity is .22 ($N=2$ validity coefficients), and the

Table 8

Validity Summary for Military Professional, Technical, and Managerial Jobs
(N = 414 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Wrist-Finger Speed</u>												
Median Validity						.07						
Weighted Mean Validity						.07						
K						2						
L						2						
M						1						
N Range						244-245(24)						
<u>Aiming</u>												
Median Validity						.13						
Weighted Mean Validity						.13						
K						3						
L						7						
M						5						
N Range						244-728(24)						
<u>Arm-Hand Steadiness</u>												
Median Validity						.06						
Weighted Mean Validity						.06						
K						1						
L						6						
M						1						
N Range						1000						
<u>Speed of Arm Movement</u>												
Median Validity						.10						
Weighted Mean Validity						.10						
K						1						
L						1						
M						1						
N Range						1194						
<u>Complex Psychomotor Predictors</u>												
Median Validity						.15	.40					
Weighted Mean Validity						.09	.36					
K						3	1					
L						47	8					
M						4	1					
N Range						11-291(139)	26-106(66)					

(Continued)

Table 8 (continued)

Validity Summary for Military Professional, Technical, and Managerial Jobs
(N = 414 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Multilimb Coordination</u>												
Median Validity	.24			.20		.28	.14					
Weighted Mean Validity	.24			.21		.27	.13					
K	1			3		14	1					
L	1			36		132	36					
M	1			4		14	3					
N Range	730			29-1000(126)		29-7093(521)	311-523(111)					
<u>Control Precision</u>												
Median Validity				.01		.22	.02					
Weighted Mean Validity				.01		.21	.06					
K				1		5	1					
L				1		20	3					
M				1		1	1					
N Range				562		90-3308(644)	331-459(103)					
<u>Race Control</u>												
Median Validity						.12	.06					
Weighted Mean Validity						.11	.06					
K						1	1					
L						6	11					
M						2	1					
N Range						307-634(407)	311					
<u>Finger Dexterity</u>												
Median Validity				.14		.08	.04					
Weighted Mean Validity				.11		.08	.05					
K				1		5	1					
L				7		33	18					
M				1		2	1					
N Range				562-1000(1000)		87-4779(797)	311-523(111)					
<u>Manual Dexterity</u>												
Median Validity						.22						
Weighted Mean Validity						.13						
K						1						
L						2						
M						1						
N Range						242-1496(849)						

median validity for aiming is .13 ($N=7$ validity coefficients). Table 8 also shows that hands-on measures of training performance are not well predicted by psychomotor test scores. The one exception to this may be complex psychomotor tests (e.g., work sample tests), which have a median correlation of .40 with hands-on training performance measures ($N=8$ validity coefficients).

Table 9 contains a summary of the validity evidence for pilots only. The great majority of these validity coefficients were obtained from military research, but the table does contain some validity data for nonmilitary pilots, also. Table 9 provides clear evidence of the validity of psychomotor tests for predicting successful completion of pilot training. Again, multilimb coordination and control precision appear to be the two abilities most directly related to pilot training performance. There is some evidence that rate control and aiming may be useful predictors as well.

Table 10 contains a summary of the validity evidence for nonmilitary professional, technical, and managerial jobs. The only abilities for which there is sufficient evidence to evaluate the probable validity are finger dexterity, manual dexterity, and wrist-finger speed. Almost all of the studies involving these abilities have been conducted with the GATB.

Table 10 suggests that all three of these abilities are moderately valid predictors of job performance for this job type. Median correlations between these abilities and subjective measures of job proficiency (e.g., supervisory ratings) are all near .15. There are not sufficient data to evaluate the correlations between these abilities and training performance. The data do suggest, however, that finger dexterity and manual dexterity are not valid predictors of educational and school achievement for this job type. The validity data for educational and school achievement are only slightly more promising for wrist-finger speed. For all three abilities, the median validity for educational and school achievement criteria is only approximately .10.

Table 11 provides a summary of the validity evidence for all professional, technical, and managerial jobs (i.e., both military and nonmilitary). Because the jobs represented in the military validity summary table for this job type differ considerably from those represented in the nonmilitary summary table, it is difficult to draw any general conclusions from Table 11. It does appear, however, that several psychomotor abilities have at least moderate validity for predicting training performance and job proficiency for these jobs. This is somewhat surprising, since it is not readily apparent how motor skills are related to training performance and job proficiency for most of the jobs in this job type. (The job of pilot is, of course, an exception.) One possible explanation may be that these validities are unrelated to psychomotor ability. In the next section, there will be a discussion of the correlations between psychomotor ability measures and cognitive ability measures. Almost all of these correlations are low to moderate and positive. Thus, for this particular job type, it may be that the low positive correlations between psychomotor abilities and criterion measures of training performance and job proficiency are attributable to cognitive ability variance embedded in the psychomotor measures.

Table 9

Validity Summary for All Pilots (N = 414 Validity Coefficients)

		CRITERION CONSTRUCT									
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
PREDICTOR CONSTRUCT											
<u>Multilimb Coordination</u>											
Median Validity			.20		.27	.12					
Weighted Mean Validity			.20		.29	.12					
K			4		14	2					
L			38		146	50					
M			6		14	4					
N Range			29-1000(56)		29-7093(521)	37-523(311)					
<u>Control Precision</u>											
Median Validity			.01		.24	.02					
Weighted Mean Validity			.01		.22	.06					
K			1		5	1					
L			1		18	3					
M			1		3	1					
N Range			562		90-1308(632)	331-459(433)					
<u>Rate Control</u>											
Median Validity					.12	.06					
Weighted Mean Validity					.11	.06					
K					1	1					
L					6	11					
M					2	1					
N Range					387-434(407)	311					
<u>Finger Dexterity</u>											
Median Validity			.14		.07	.04					
Weighted Mean Validity			.11		.08	.03					
K			7		5	1					
L			7		29	18					
M			1		2	1					
N Range			562-1000(1000)		87-4779(894)	311-523(111)					
<u>Manual Dexterity</u>											
Median Validity			.09		.09						
Weighted Mean Validity			.09		.1						
K			1		1						
L			1		1						
M			1		1						
N Range			1496								

(Continued)

Table 9 (continued)

Validity Summary for All Pilots (N = 414 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement	Training Performance					Job Proficiency			Job Involvement/Withdrawal			
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
PREDICTOR CONSTRUCT												
<u>Wrist-Finger Speed</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Aiming</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Arm-Hand Steadiness</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Speed of Arm Movement</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Complex Psychomotor Predictors</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												

Table 10

**Validity Summary for Nonmilitary Professional, Technical, and Managerial Jobs
(N = 379 Validity Coefficients)**

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Multitask Coordination</u>												
Median Validity				.06			-.05					
Weighted Mean Validity				.06			-.00					
K				1			1					
L				4			14					
M				2			2					
N Range				37			37					
<u>Control Precision</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Rate Control</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Finger Dexterity</u>												
Median Validity	.10	.12						.13	.35			
Weighted Mean Validity	.10	.09				.16		.12	.35			
K	1	1				1		1	1			
L	52	5				1		46	1			
M	1	1				1		1	1			
N Range	30-271(64)	50-130(78)				80		26-232(60)	41			
<u>Manual Dexterity</u>												
Median Validity	.06	.11				.20		.16	.30			
Weighted Mean Validity	.06	.14				.20		.18	.30			
K	1	1				1		1	1			
L	52	5				1		46	1			
M	1	1				1		1	1			
N Range	30-271(64)	50-130(78)				80		26-232(60)	41			

(Continued)

Validity Summary for Nonmilitary Professional, Technical, and Managerial Jobs
(N = 379 validity coefficients)

70

Table 11

Validity Summary for All Professional, Technical, and Managerial Jobs
(N = 793 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	
<u>Multilimb Coordination</u>											
Median Validity	.24			.20		.26	.12				
Weighted Mean Validity	.24			.20		.28	.12				
K	1			4		14	2				
L	1			38		152	50				
M	1			6		14	4				
N Range	730			29-1000(56)		29-7093(523)	37-523(311)				
<u>Control Precision</u>											
Median Validity				.01		.22	.02				
Weighted Mean Validity				.01		.21	.06				
K				1		5	1				
L				1		20	3				
M				1		3	1				
N Range				562		30-3308(644)	331-459(403)				
<u>Rate Control</u>											
Median Validity						.12	.06				
Weighted Mean Validity						.11	.06				
K						1	1				
L						6	11				
M						2	1				
N Range						387-434(407)	311				
<u>Finger Dexterity</u>											
Median Validity	.10	.12		.14		.08	.04	.13	.35		
Weighted Mean Validity	.10	.09		.11		.08	.05	.12	.35		
K	1	1		1		6	1	1	1		
L	52	5		7		34	18	46	1		
M	1	1		1		3	1	1	1		
N Range	30-271(64)	50-130(78)		562-1000(1000)		80-4779(764)	311-523(311)	26-232(60)	41		
<u>Manual Dexterity</u>											
Median Validity	.06	.11				.20		.16	.30		
Weighted Mean Validity	.06	.14				.13		.15	.30		
K	1	1				2		1	1		
L	52	5				3		46	1		
M	1	1				2		1	1		
N Range	30-271(64)	50-130(78)				80-1496(242)		26-232(60)	41		

(Continued)

Table 11 (continued)

Validity Summary for All Professional, Technical, and Managerial Jobs
(N = 793 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT												
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal		
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
<u>Wrist-Finger Speed</u>													
Median Validity	.10	.19						.15	.14				
Weighted Mean Validity	.11	.20						.15	.14				
K	1	1						1	1				
L	95	5						47	1				
M	3	1						1	1				
N Range	30-1079(81)	50-130(78)				90-245(244)		30-232(60)	41				
<u>Aiming</u>													
Median Validity						.13							
Weighted Mean Validity						.13							
K						3							
L						7							
M						5							
N Range						244-728(244)							
<u>Arm-Hand Steadiness</u>													
Median Validity													
Weighted Mean Validity						.09							
K						.08							
L						3							
M						19							
N Range						6							
<u>Speed of Arm Movement</u>													
Median Validity													
Weighted Mean Validity						.10							
K						.10							
L						1							
M						.1							
N Range						1194							
<u>Complex Psychomotor Predictors</u>													
Median Validity													
Weighted Mean Validity						.15							
K						.09							
L						3							
M						47							
N Range						107							
						43-293(139)							
						26-104(15)							

Table 12

Validity Summary for Nonmilitary Clerical Jobs (N = 143 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal		
		Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
PREDICTOR CONSTRUCT												
<u>Multitask Coordination</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
Control Precision												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
Rate Control												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
Finger Dexterity												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
31-322(62) 50-130(59) 50-60(55)												
Manual Dexterity												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
31-322(62) 50-130(59) 50-60(55)												

(Continued)

Table 12 (continued)

Validity Summary for Nonmilitary Clerical Jobs (N = 143 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Wrist-Finger Speed</u>												
Median Validity	.25	.38		.25				.10	.26	.20		
Weighted Mean Validity	.25	.38		.25				.11	.22	.19		
K	1	1		1				2	1	1		
L	1	1		1				38	6	2		
M	1	1		1				1	1	1		
N Range	51	66		59				31-322(64)	50-130(59)	50-60(55)		
<u>Aiming</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Arm-Hand Steadiness</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Speed of Arm Movement</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Complex Psychomotor Predictors</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												

Clerical Jobs. No validity data were identified for clerical jobs in the military. Therefore Table 12, which summarizes the validity data for nonmilitary clerical jobs, is the only summary table for this job type.

Validity data are only available for three abilities for this job type: finger dexterity, manual dexterity, and wrist-finger speed. In fact, all of the validity data summarized in Table 12 are from studies involving the GATB psychomotor scales.

All three abilities appear to be relatively valid predictors for nonmilitary clerical jobs. For example, most of the median correlations between these abilities and job proficiency criteria fall between .10 and .20. Validity coefficients for educational and school achievement criteria are somewhat higher. All but one of the median validities in these six predictor-criterion cells are between .25 and .38. Unfortunately, there are not sufficient data to evaluate the criterion-related validity of these abilities for training performance or job involvement/withdrawal criteria.

Information in Table 12 also reveals that the validities for wrist-finger speed tend to be slightly higher than the validities for finger and manual dexterity across most of the criterion constructs for which validity data are available. While the validity differences between wrist-finger speed and the other two abilities are not overwhelming, they do suggest that wrist-finger speed might be the first choice for a psychomotor predictor for this job type.

Sales Jobs. Since there are no data for sales jobs in the military (see Table 6), validity data were summarized for nonmilitary sales jobs only. Table 13 contains this validity summary.

In total, only 15 validity coefficients were identified for this job type. These validity coefficients were obtained from five studies in which the GATB was used to predict supervisory ratings of job proficiency. As Table 13 shows, the median validity coefficients across these five studies are very small for all three psychomotor abilities assessed by the GATB. The median validity for wrist-finger speed is only .14, and the median validities for finger dexterity and manual dexterity are both only .08.

These data indicate that psychomotor abilities add little to the prediction of job proficiency for sales jobs. It must be recalled, of course, that this conclusion is based on very limited data. Nevertheless, it seems unlikely that many sales positions would require extensive psychomotor skills. Indeed, the lack of psychomotor validity data probably reflects the belief of many psychologists that psychomotor predictors do not warrant investigation for this job type. The sketchy data available from the GATB seem to support this belief.

Protective Service Jobs. Table 14 summarizes the validity data for military protective service jobs. Most of the validity coefficients in this table are the result of the AAF's World War II aircrew selection research on bombardiers (Melton, 1947). The validity coefficients for complex psychomotor predictors were taken from recent research into the selection of tank crewmen (Campbell & Black, 1982; Eaton et al., 1980).

Table 13

Validity Summary for Nonmilitary Sales Jobs (N = 15 Validity Coefficients)

CRITERION CONSTRUCT											
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
PREDICTOR CONSTRUCT											
Multilimb Coordination											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
Control Precision											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
Rate Control											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
Finger Dexterity											
Median Validity											
Weighted Mean Validity							.08				
K							.09				
L							2				
H							5				
N Range							1				
							52-133 (59)				
Manual Dexterity											
Median Validity							.08				
Weighted Mean Validity							.11				
K							2				
L							5				
H							1				
N Range							52-133 (59)				

(Continued)

Table 13 (continued)

Validity Summary for Nonmilitary Sales Jobs (N = 15 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Go-No Go Training Courses	Man-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Wrist-Finger Speed</u>								.14				
Median Validity								.16				
Weighted Mean Validity								2				
K								5				
L								1				
M								52-133 (59)				
N Range												
<u>Aiming</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Arm-Hand Steadiness</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Speed of Arm Movement</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Complex Psychomotor Predictors</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												

Table 14

Validity Summary for Military Protective Service Jobs (N = 118 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal		
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
PREDICTOR CONSTRUCT												
<u>Multilimb Coordination</u>												
Median Validity												
Weighted Mean Validity												
K			.14		.20	.23	.15	.14				
L			.14		.16	.15	.18	.15				
M			1		1	1	1	1				
			1		23	14	9	2				
			1		4	5	1	2				
N Range			574		30-1829(363)	32-574(164)	33-165(145)	27-32(30)				
<u>Control Precision</u>												
Median Validity												
Weighted Mean Validity												
K					.14							
L					.13							
M					1							
					7							
N Range					2							
<u>Rate Control</u>												
Median Validity												
Weighted Mean Validity												
K						.05						
L						.05						
M						1						
						4						
N Range						1						
						164						
<u>Finger Dexterity</u>												
Median Validity												
Weighted Mean Validity												
K					.12	.06		-.32				
L					.13	.08		-.32				
M					1	1		1				
					11	5		1				
N Range					2	1		1				
					126-1828(363)	164-486(195)		32				
<u>Manual Dexterity</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												

(Continued)

Table 14 (continued)

Validity Summary for Military Protective Service Jobs (N = 118 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance				Job Proficiency				Job Involvement/ Withdrawal
	Course Grades	Instructor Evaluations	Objective Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	
Wrist-Finger Speed Median Validity Weighted Mean Validity K L H N Range											
Aiming Median Validity Weighted Mean Validity K L H N Range											
Arm-Hand Steadiness Median Validity Weighted Mean Validity K L H N Range											
Speed of Arm Movement Median Validity Weighted Mean Validity K L H N Range											
Complex Psychomotor Predictors Median Validity Weighted Mean Validity K L H N Range											

Table 14 shows that multilimb coordination, control precision, and finger dexterity measures are all moderately valid predictors of graduation from bombardier training. Median r 's range from .12 to .20. Validity coefficients were somewhat lower when hands-on training measures were used as criteria. For bombardiers, the most common hands-on measure was the distance between the bomb target and the actual spot where the bombs were dropped. The median correlation between hands-on training measures and measures of multilimb coordination was .23 ($N=14$ validity coefficients), while median validities for rate control, finger dexterity, and arm-hand steadiness measures using this same criterion were .05 ($N=4$ validity coefficients), .06 ($N=5$ validity coefficients), and .06 ($N=3$ validity coefficients), respectively.

Thirty-eight validity coefficients were tabulated from research on the prediction of tank crew training performance using complex psychomotor predictors. The median value of these 38 validity coefficients is only .07. The validities vary greatly from test to test and from criterion measure to criterion measure, however. For example, the Gunner Tracking Task used by Campbell and Black (1982) had moderately high predictive validity (median $r=.17$, $N=2$ validity coefficients). Additional research may show that certain complex psychomotor predictors are quite useful for predicting success in military protective service jobs.

Table 15 contains a summary of the validity data for nonmilitary protective service jobs. Again, very few validity coefficients are available. Only five criterion-related validity studies were identified for this job family--all involving the GATB.

Table 15 indicates that wrist-finger speed is a moderately valid predictor of subjective measures of job proficiency (median $r=.23$, $N=5$ validity coefficients). Median validities are somewhat lower for finger and manual dexterity ($r=.10$ and $.15$, respectively, $N=5$ validity coefficients each). While there are too few coefficients to draw any definitive conclusions, the data do suggest that all three of these psychomotor abilities may have moderate validity for predicting supervisory job proficiency ratings.

Table 16 provides a complete summary of the validity data for this job type (i.e., combining military and nonmilitary research). Based on this table, it appears that multilimb coordination and wrist-finger speed are the two psychomotor abilities with the highest criterion-related validities for protective service jobs. Median validities for these two abilities across all criteria are near .20. In addition, control precision, finger dexterity, and manual dexterity appear to have moderate criterion-related validity (i.e., $.10 < r < .20$) for this job type.

In sum, even though relatively few criterion-related validity coefficients were identified for protective service jobs, those data which have been collected suggest that psychomotor abilities may indeed be useful for predicting training performance and job proficiency for these jobs.

Service Jobs. Tables 17 and 18 contain summaries of the criterion-related validity evidence for military and nonmilitary service jobs, respectively.

Table 15

Validity Summary for Nonmilitary Protective Service Jobs (N = 15 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Multilimb Coordination</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Control Precision</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Rate Control</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Finger Dexterity</u>												
Median Validity								.10				
Weighted Mean Validity								.14				
K								1				
L								5				
M								1				
N Range								51-166(60)				
<u>Manual Dexterity</u>												
Median Validity								.15				
Weighted Mean Validity								.10				
K								1				
L								5				
M								1				
N Range								51-166(60)				

(Continued)

Table 15 (continued)

Validity Summary for Nonmilitary Protective Service Jobs (N = 15 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Wrist-Finger Speed</u> Median Validity Weighted Mean Validity K L H N Range								.23 .15 1 5 1 51-166(60)				
<u>Aiming</u> Median Validity Weighted Mean Validity K L H N Range												
<u>Arm-Hand Steadiness</u> Median Validity Weighted Mean Validity K L H N Range												
<u>Speed of Arm Movement</u> Median Validity Weighted Mean Validity K L H N Range												
<u>Complex Psychomotor Predictors</u> Median Validity Weighted Mean Validity K L H N Range												

Table 16

Validity Summary for All Protective Service Jobs (N = 133 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
			Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Man-is-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
Multitask Coordination												
Median Validity				.14		.20	.23	.15	.14			
Weighted Mean Validity				.14		.16	.15	.18	.15			
K				1		1	1	1	1			
L				1		23	14	9	2			
M				1		4	5	1	2			
N Range				574		30-1829(36)	32-574(364)	33-165(165)	27-32(30)			
Control Precision												
Median Validity						.14						
Weighted Mean Validity						.13						
K						1						
L						7						
M						2						
N Range						74-1829(330)						
Rate Control												
Median Validity							.05					
Weighted Mean Validity							.05					
K							1					
L							4					
M							1					
N Range							164					
Finger Dexterity												
Median Validity						.13	.06	.10	-.32			
Weighted Mean Validity						.13	.08	.14	-.32			
K						1	1	1	1			
L						11	5	5	1			
M						2	2	1	1			
N Range						126-1828(3)	37164-486(105)	51-166(60)	32			
Manual Dexterity												
Median Validity								.15				
Weighted Mean Validity								.10				
K								1				
L								5				
M								1				
N Range								51-166 (60)				

(Continued)

Table 16 (continued)
 Validity Summary for All Protective Service Jobs (N = 133 Validity Coefficients)

CRITERION CONSTRUCT												
PREDICTOR CONSTRUCT	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Wrist-Finger Speed</u> Median Validity Weighted Mean Validity K L H N Range								.23 .15 1 5 1 51-166 (50)				
<u>Aiming</u> Median Validity Weighted Mean Validity K L H N Range												
<u>Arm-Hand Steadiness</u> Median Validity Weighted Mean Validity K L H N Range												
<u>Speed of Arm Movement</u> Median Validity Weighted Mean Validity K L H N Range								.06 .06 1 3 1 86-144(108)				
<u>Complex Psychomotor Predictors</u> Median Validity Weighted Mean Validity K L H N Range												
								.07 .05 2 38 4 16-147(26)				

Table 17

Validity Summary for Military Service Jobs (N = 8 Validity Coefficients)

	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
			Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures		
PREDICTOR CONSTRUCT	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Multilimb Coordination</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
<u>Control Precision</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
<u>Rate Control</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
<u>Finger Dexterity</u>												
Median Validity				.12	-.12							
Weighted Mean Validity				.12	-.12							
K				1	1							
L				2	2							
H				1	1							
N Range				51	38							
<u>Manual Dexterity</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												

(Continued)

Table 17 (continued)

Validity Summary for Military Service Jobs (N = 8 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	
<u>Wrist-Finger Speed</u>											
Median Validity				.10	-.19						
Weighted Mean Validity				.10	-.19						
K				1	1						
L				1	1						
M				1	1						
N Range				79	62						
<u>Aiming</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Arm-Hand Steadiness</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Speed of Arm Movement</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Complex Psychomotor Predictors</u>											
Median Validity				-.12	.17						
Weighted Mean Validity				-.12	.17						
K				1	1						
L				1	1						
M				1	1						
N Range				54	48						

Table 18

Validity Summary for Nonmilitary Service Jobs (N = 162 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance				Job Proficiency				Job Involvement/Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	
<u>Multilimb Coordination</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Control Precision</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Rate Control</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Finger Dexterity</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Manual Dexterity</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											

(Continued)

Table 18 (continued)

Validity Summary for Nonmilitary Service Jobs (N = 162 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
Wrist-Finger Speed												
Median Validity	.21	.25		.29			.02	.20				
Weighted Mean Validity	.18	.21		.29			.02	.17				
K	1	1		1			1	2				
L	5	7		1			1	33				
M		1		1			1	1				
N Range	65-160(23)	51-99(61)		66			165	31-199(60)				
Aiming												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Arm-Hand Steadiness												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Speed of Arm Movement												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Complex Psychomotor Predictors												
Median Validity									.00	.02		
Weighted Mean Validity									-.01	.03		
K									1	1		
L									3	6		
M									3	3		
N Range									100-103(360)	100-103(360)		

In total, only eight validity coefficients were identified for military service jobs. This is insufficient to draw any conclusions regarding the criterion-related validity of psychomotor tests for these jobs.

The validity coefficients in Table 18 were derived from two sources. The first was a concurrent validity study of Washington, DC taxicab drivers reported by Farr, O'Leary, Pfeiffer, Goldstein, and Bartlett (1971). All of the validities for job-related and archival measures of job proficiency were taken from this study. The remaining validity coefficients were based on research with the GATB psychomotor scales.

Median validities from the taxicab study are all quite low, ranging from $-.16$ to $.10$. These data suggest that multilimb coordination, rate control, and complex psychomotor predictors (in this case, driving simulators) are unrelated to job proficiency criteria. It should be noted, however, that the reliability of job-related and archival measures of job proficiency tends to be fairly low. An examination of all of the validity summary tables which were prepared for this section indicated that validities tended to be lower for these two criteria than they were for other educational and school achievement, training performance, or job proficiency criteria. Thus, while the validities for taxicab drivers in Table 18 are certainly not encouraging, it may be that the low correlations reflect problems in the selection and measurement of the criterion rather than a lack of criterion-related validity.

In contrast, validity coefficients based on GATB research indicate that finger dexterity, manual dexterity, and wrist-finger speed are all valid predictors of several different criteria. For example, median validities for instructor evaluations range from $.25$ to $.42$. Median ability correlations with subjective measures of training performance range from $.29$ to $.50$. Finally, the criterion-related validity for subjective measures of job proficiency ranges from $.13$ to $.20$ for these three abilities.

In general, comparing across criteria, there appears to be little difference in the criterion-related validity of these three abilities. Comparing across abilities, subjective criterion measures for several different criterion categories (e.g., instructor evaluations of educational and school achievement, subjective measures of training performance, and subjective measures of job proficiency) appear to be predictable from psychomotor ability tests. Thus, for this job type, psychomotor abilities may be useful for predicting both training and job proficiency criteria.

Table 19 summarizes the validity data for all service jobs. Since military research has contributed very little validity data for this job type, combining military and nonmilitary validity data did not result in any additional conclusions beyond those already described above for nonmilitary service jobs.

Mechanical and Structural Maintenance Jobs. No military criterion-related validity research was located for mechanical and structural maintenance jobs. Therefore, the validity summary for nonmilitary mechanical and structural jobs in Table 20 contains all the validity coefficients which were identified for this job type.

Validity Summary for All Service Jobs (N = 170 Validity Coefficients)

(Continued)

Table 19 (continued)

Validity Summary for All Service Jobs (N = 170 Validity Coefficients)

CRITERION CONSTRUCT												
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
PREDICTOR CONSTRUCT												
Wrist-Finger Speed												
Median Validity	.21	.25		.20	-.19		.02	.20				
Weighted Mean Validity	.18	.21		.17	-.19		.02	.18				
K	1	1		2	1		1	2				
L	5	7		2	1		1	33				
M	1	1		2	1		1	1				
N Range	65-160(78)	51-99(61)		46-79(62)	62		145	11-199(60)				
Aiming												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Arm-Hand Steadiness												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Speed of Arm Movement												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Complex Psychomotor Predictors												
Median Validity				-.12	.17				.00	.02		
Weighted Mean Validity				-.12	.17				-.01	.03		
K				1	1				1	1		
L				1	1				3	6		
M				1	1				3	6		
N Range				54	48				100-303(300)	100-303(300)		

Table 20

Validity Summary for Nonmilitary Mechanical and Structural Maintenance Jobs
(N = 184 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											Job Involvement/ Withdrawal	
	Educational and School Achievement		Training Performance					Job Proficiency			Job Satisfaction		Job Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job- Related Measures	Archival Measures			
Multilimb Coordination													
Median Validity													
Weighted Mean Validity													
K													
L													
M													
N Range													
Control Precision													
Median Validity													
Weighted Mean Validity													
K													
L													
M													
N Range													
Rate Control													
Median Validity													
Weighted Mean Validity													
K													
L													
M													
N Range													
Finger Dexterity													
Median Validity	.16	.16	.28	.26			.22	.21					
Weighted Mean Validity	.18	.18	.28	.23			.22	.17					
K	1	1	1	1			1	2					
L	10	8	1	3			2	39					
M	1	1	1	1			2	2					
N Range	11-75(57)	40-702(60)	75	62-86(83)			430	31-322(54)					
Manual Dexterity													
Median Validity	.20	.18	.21	.18				.21					
Weighted Mean Validity	.13	.20	.21	.27				.21					
K	1	1	1	1				2					
L	10	8	1	3				40					
M	1	1	1	1				2					
N Range	11-75(57)	40-702(60)	75	62-86(83)				31-322(54)					

(Continued)

Table 20 (continued)

Validity Summary for Nonmilitary Mechanical and Structural Maintenance Jobs
(N = 184 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Wrist-Finger Speed</u>												
Median Validity	.20	.18		.06				.16				
Weighted Mean Validity	.20	.17		.18				.18				
K	1	1		1				1				
L	10	8		3				.38				
M	1	1		1				.1				
N Range	33-75(5)	40-202(60)		62-86(83)				31-322(54)				
<u>Aiming</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Arm-Hand Steadiness</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Speed of Arm Movement</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
<u>Complex Psychomotor Predictors</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												

The vast majority of the validity data summarized in Table 20 are based on research with the GATB psychomotor scales. Several other psychomotor tests and studies are also represented in this table, however.

Sufficient validity data are available to evaluate the criterion-related validity of only three psychomotor abilities: finger dexterity, manual dexterity, and wrist-finger speed. All three abilities appear to have moderate validity for predicting educational and school achievement. Median validities for these three abilities across the two educational and school achievement criteria range from .16 to .20. The few validities available for training criteria suggest that training performance may be even more predictable than educational and school achievement for this job type. Median validities for finger dexterity and manual dexterity range from .18 to .28. Correlations between wrist-finger speed and training performance are somewhat lower, however (median $r=.06$, $N=3$ validity coefficients). Finally, there is little difference in the median validities of these three abilities for subjective measures of job proficiency. Median correlations are .21, .21, and .14 for finger dexterity, manual dexterity, and wrist-finger speed, respectively ($N=39$, 40, and 38 validity coefficients, respectively).

In summary, criterion-related validity research shows that finger dexterity, manual dexterity, and wrist-finger speed are all equally valid predictors of educational and school achievement and subjective measures of job proficiency for mechanical and structural maintenance jobs. Median validity coefficients for all three abilities for these criteria tend to range from .15 to .20. The validity coefficients available for training performance and job proficiency suggest that manual dexterity and finger dexterity might be more valid predictors than wrist-finger speed.

Electronics Jobs. Table 21 summarizes the few validity coefficients identified for electronics jobs. All of the psychomotor criterion-related validity research identified for this job type was conducted with non-military subjects. The only tests used in this research were the GATB Finger Dexterity, Manual Dexterity, and Motor Coordination scales. In total, only 30 validity coefficients were located, 10 for each of the three GATB scales.

The validity data which do exist suggest that finger dexterity (median $r=.17$, $N=3$ validity coefficients) and wrist-finger speed (median $r=.12$, $N=3$ validity coefficients) are moderately valid predictors of school grades for this job type. Median validities for subjective measures of job proficiency are comparable for the three abilities. The median validity for finger dexterity is .15, the median validity for manual dexterity is .12, and the median validity for wrist-finger speed is .16 ($N=7$ validity coefficients each).

Based on the validity data summarized in Table 21, it is difficult to determine if any of the four major criterion categories are more predictable than others. Table 21 does suggest, however, that the criterion-related validity for manual dexterity is lower than for finger dexterity or wrist-finger speed.

Table 21

Validity Summary for Nonmilitary Electronics Jobs (N = 30 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	
<u>Multitimb Coordination</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Control Precision</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Rate Control</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
H											
N Range											
<u>Finger Dexterity</u>											
Median Validity											
Weighted Mean Validity											
K	.17							.13			
L	.13							.16			
H	1							1			
N Range	3							7			
Median Validity	1							1			
Weighted Mean Validity	58-97(62)							50-124(59)			
K											
L											
H											
N Range											
<u>Manual Dexterity</u>											
Median Validity											
Weighted Mean Validity											
K	.04							.12			
L	.02							.11			
H	1							1			
N Range	3							7			
Median Validity	1							1			
Weighted Mean Validity	58-97(62)							50-124(59)			
K											
L											
H											
N Range											

(Continued)

Table 21 (continued)

Validity Summary for Nonmilitary Electronics Jobs (N = 30 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement	Training Performance						Job Proficiency			Job Involvement/Withdrawal		
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>PREDICTOR CONSTRUCT</u> <u>Wrist-Finger Speed</u> Median Validity Weighted Mean Validity K L M N Range	.12							.16				
	.17							.17				
	1							1				
	3							7				
	1							1				
	58-97(62)							60-124(59)				
<u>Aiming</u> Median Validity Weighted Mean Validity K L M N Range												
<u>Arm-Hand Steadiness</u> Median Validity Weighted Mean Validity K L M N Range												
<u>Speed of Arm Movement</u> Median Validity Weighted Mean Validity K L M N Range												
<u>Complex Psychomotor Predictors</u> Median Validity Weighted Mean Validity K L M N Range												

Industrial Jobs. All of the validity coefficients for industrial jobs are summarized in Table 22. Since there are no data on industrial jobs in the military (see Table 6), Table 22 includes only nonmilitary research.

As with most of the other nonmilitary validity summary tables described thus far, Table 22 is dominated by validity coefficients obtained from GATB research. For example, except for two studies involving complex psychomotor predictors, the only predictors used in the validity research summarized in Table 22 have been measures of finger dexterity, manual dexterity, or wrist-finger speed. Some of the median validities reported in Table 22, however, are based on research with a number of different finger dexterity or manual dexterity tests (see the *M* entry in the summary table). This should mean that the conclusions below are general to validity results which would be obtained across a wide variety of tests of finger and manual dexterity.

Results for educational and school achievement indicate that both finger and manual dexterity are valid predictors. Median validities for these two abilities across the two educational and school achievement criteria range from .22 to .31. The median validities for wrist-finger speed are somewhat lower, however ($r=.08$ with school grades and .18 with instructor evaluations, $N=4$ validity coefficients each).

Results are very similar for subjective measures of training performance. Both finger dexterity and manual dexterity possess moderately high criterion-related validity, while the criterion-related validity of wrist-finger speed is somewhat lower. In studies involving hands-on measures of training performance, however, neither finger nor manual dexterity was significantly related to the criterion measure.

Validity results varied significantly across the three job proficiency criteria. Median validities for all three abilities for subjective measures of job proficiency were near .25. Median validities for archival job proficiency measures were slightly lower, ranging from .17 to .21. Median validities for job-related measures of job proficiency (i.e., hands-on job performance tests) were substantially lower than the median validities for the other two job proficiency criteria, ranging from .00 for finger dexterity to .14 for manual dexterity and wrist-finger speed ($N=1$ validity coefficient each). As noted previously in the validity summary for service jobs, throughout the validity summary tables median validities for job-related and archival measures of job proficiency tend to be lower than the median validities for other criteria. This suggests that there may be reliability problems associated with the measurement of these two job proficiency criteria.

Finally, Table 22 suggests that job withdrawal (i.e., turnover and absenteeism) is not readily predictable from manual dexterity for this job type (median $r=.02$, $N=3$ validity coefficients). The only validity coefficient identified for job withdrawal and finger dexterity is somewhat higher ($r=.14$), suggesting that there may be some value to investigating this psychomotor ability as a predictor of job withdrawal for industrial jobs.

A comparison of the criterion-related validity of finger dexterity, manual dexterity, and wrist-finger speed indicates that the median

Validity Summary for Nonmilitary Industrial Jobs (N = 814 Validity Coefficients)

(Continued)

Table 22 (continued)

Validity Summary for Nonmilitary Industrial Jobs (N = 814 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal		
			Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>PREDICTOR CONSTRUCT</u> <u>Wrist-Finger Speed</u> Median Validity Weighted Mean Validity K L M N Range	.08	.18		.16	.30		.12	.22	.14	.21		
	.15	.24		.16	.30		.12	.22	.14	.23		
	1	1		1	1		1	2	1	1		
	4	4		2	1		1	226	1	17		
	1	1		1	1		1	1	1	1		
	50-79(57)	51-164(57)		53-64(58)	81		49	30-183(55)	57	32-100(52)		
<u>Aiming</u> Median Validity Weighted Mean Validity K L M N Range												
<u>Arm-Hand Steadiness</u> Median Validity Weighted Mean Validity K L M N Range												
<u>Speed of Arm Movement</u> Median Validity Weighted Mean Validity K L M N Range												
<u>Complex Psychomotor Predictors</u> Median Validity Weighted Mean Validity K L M N Range							.14					-.09
							.14					-.09
							1					1
							2					1
							1					1
							38					63

validities for wrist-finger speed are somewhat lower than those for finger dexterity and manual dexterity. This is especially evident for educational and school achievement and training performance criteria. In addition, for these two criterion categories, the criterion-related validity of finger dexterity tends to be slightly higher than the criterion-related validity of manual dexterity. For job proficiency criteria, the median criterion-related validities for all three abilities are virtually identical.

Thus, the validity coefficients summarized in Table 22 once again illustrate the criterion-related validity of psychomotor abilities for predicting educational and school achievement, training performance, and job proficiency. For two of these criterion categories--educational and school achievement and training performance--the criterion-related validities for finger dexterity and manual dexterity are higher than those for wrist-finger speed. Of special interest in Table 22 is one study which indicates that finger dexterity might be a valid predictor of job withdrawal. Moreover, when there are differences in the median validities for finger dexterity and manual dexterity, those for finger dexterity tend to be slightly higher. These findings suggest that finger dexterity might be the single best psychomotor ability for predicting the various criteria for this job type.

Miscellaneous Jobs. Table 23 summarizes the criterion-related validity data for military miscellaneous jobs. Only seven validity coefficients were identified for this table. This is insufficient to draw any general conclusions about the validity of psychomotor tests for these jobs.

Table 24 contains a summary of the validity evidence for nonmilitary miscellaneous jobs. Jobs classified into this job type included air traffic controller, college student, dairy farm hand, lemon picker, order filler, and power plant operator.

Virtually all of the 84 validity coefficients summarized in Table 24 were taken from GATB research. Only one of the validity coefficients included in this table was not based on one of the GATB psychomotor scales. Thus, Table 24 summarizes the criterion-related validity evidence for a wide variety of jobs using only a very few psychomotor tests.

These problems--especially the problems associated with the wide variety of jobs--make it difficult to interpret the median validities shown in the table. For example, the median correlation between manual dexterity and school grades is $-.02$, but the median correlation between manual dexterity and subjective measures of job proficiency is $.24$. How are these results to be reconciled? One plausible explanation is that the jobs for which school grades were used as a criterion differ substantially from those for which subjective job proficiency measures were used as a criterion. In this case, that may indeed be a correct explanation. For the 18 validity observations for which school grades were used as a criterion, the "job" was most often college student with major field of study unspecified. In contrast, one of the jobs for which subjective measures of job proficiency was used as the criterion was lemon picker. Since the psychomotor abilities relevant to success in these two "jobs" seem to overlap little if at all, it is not surprising that there is such a large difference between the median validities in these two cells of the table.

Table 23

Validity Summary for Military Miscellaneous Jobs (N = 7 Validity Coefficients)

	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
			Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures		
PREDICTOR CONSTRUCT	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
Multilimb Coordination												
Median Validity				.31	.04							
Weighted Mean Validity				.31	.04							
K				1	1							
L				2	2							
M				2	2							
N Range				52-69 (60)	381							
Control Precision												
Median Validity				.02								
Weighted Mean Validity				.02								
K				1	1							
L				1	1							
M				1	1							
N Range				47								
Rate Control												
Median Validity					.02							
Weighted Mean Validity					.02							
K					1							
L					1							
M					1							
N Range					381							
Finger Dexterity												
Median Validity				.19								
Weighted Mean Validity				.19								
K				1	1							
L				1	1							
M				1	1							
N Range				69								
Manual Dexterity												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												

(Continued)

Table 23 (continued)
Validity Summary for Military Miscellaneous Jobs (N = 7 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT												
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal		
			Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures			Archival Measures
Wrist-Finger Speed Median Validity Weighted Mean Validity K L H N Range													
Aiming Median Validity Weighted Mean Validity K L H N Range													
Arm-Hand Steadiness Median Validity Weighted Mean Validity K L H N Range													
Speed of Arm Movement Median Validity Weighted Mean Validity K L H N Range													
Complex Psychomotor Predictors Median Validity Weighted Mean Validity K L H N Range													

Table 24

Validity Summary for Nonmilitary Miscellaneous Jobs (N = 84 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal		
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
PREDICTOR CONSTRUCT												
<u>Multilimb Coordination</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
<u>Control Precision</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
<u>Rate Control</u>												
Median Validity												
Weighted Mean Validity												
K												
L												
H												
N Range												
<u>Finger Dexterity</u>												
Median Validity												
Weighted Mean Validity												
K	.16						.20					
L	.12			.32			.17					
H	2			1			1					
L	18			1			8					
H	1			1			1					
N Range												
50-234(72)												
50-152(86)												
<u>Manual Dexterity</u>												
Median Validity												
Weighted Mean Validity												
K	-.02			.36			.24					
L	-.03			.36			.25					
H	2			1			2					
L	18			1			9					
H	1			1			1					
N Range												
50-234(72)												
50-212(86)												

(Continued)

Table 24 (continued)

Validity Summary for Nonmilitary Miscellaneous Jobs (N = 84 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	
<u>Wrist-Finger Speed</u>											
Median Validity	.16				.34			.30			
Weighted Mean Validity	.17				.34			.24			
K	2				1			2			
L	18				1			9			
M	1				1			1			
N Range	50-234(72)				54			50-212(86)			
<u>Aiming</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Arm-Hand Steadiness</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Speed of Arm Movement</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Complex Psychomotor Predictors</u>											
Median Validity											
Weighted Mean Validity											
K							.75				
L											
M											
N Range											

In general, in spite of these problems, the median validities reported in Table 24 are moderate to high. Excluding the manual dexterity-school grades cell and all predictor-criterion cells with only one validity observation, median validities range from .16 to .24. Median validities are slightly higher for wrist-finger speed than for finger or manual dexterity.

In summary, Table 24 demonstrates that psychomotor abilities are related to educational and job proficiency criteria for a number of miscellaneous jobs. Because of the range of jobs included in Table 24, however, it is difficult to draw any specific conclusions about the criterion-related validity of any particular psychomotor ability for any specific class of miscellaneous jobs. Thus, the conclusions which can be drawn based on this table are quite limited.

Table 25 contains a summary of the validity data for all miscellaneous jobs (i.e., both military and nonmilitary). Since so few validity coefficients were identified for military jobs, the conclusions which can be drawn from Table 25 do not differ from those summarized above for Table 24.

Validity Results by Research Setting. Tables 26 and 27 provide over-all summaries of the validity data for military and nonmilitary research, respectively.

Table 26 shows that most psychomotor research in the military has focused on coordination and tracking abilities, such as multilimb coordination, control precision, rate control, aiming, and arm-hand steadiness. By contrast, there has been relatively little research into the criterion-related validity of dexterity abilities such as manual dexterity and wrist-finger speed. There have been a number of studies, however, investigating the criterion-related validity of finger dexterity.

The validity results for military research show that median validities for multilimb coordination tend to be higher than those for any other psychomotor ability. These median validities average more than .20 across several different training performance criteria. Other abilities with moderate to high median validities include control precision (median $r=.19$ with graduation from training, $N=27$ validity coefficients), aiming (median $r=.13$ with graduation from training, $N=7$ validity coefficients), and rate control (median $r=.12$ with graduation from training, $N=6$ validity coefficients). Median validities for finger dexterity, wrist-finger speed, and arm-hand steadiness average .10 or less across criteria, indicating that these abilities have little or no criterion-related validity for the military jobs studied. Validity results for studies using complex psychomotor predictors vary greatly. Median validities are .15 for graduation from training ($N=47$ validity coefficients) and .10 for hands-on training measures ($N=46$ validity coefficients), indicating that complex psychomotor predictors were only moderately valid on the average. There has been insufficient research to draw any conclusions regarding the criterion-related validity of manual dexterity and speed of arm movement.

The validity results summarized in Table 26 are based on a very unrepresentative sample of military jobs. The vast majority of these validity coefficients are based on research with pilots ($N=396$ validity coefficients, or 72.4%). None of the studies reviewed for Table 26 involved

Table 25

Validity summary for All Miscellaneous Jobs (N = 91 Validity coefficients)

	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
			Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures		
PREDICTOR CONSTRUCT	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
<u>Multilimb Coordination</u>												
Median Validity				.31	.04							
Weighted Mean Validity				.31	.04							
K				1	1							
L				2	2							
M				2	2							
N Range				52-69(60)	381							
<u>Control Precision</u>												
Median Validity				.02								
Weighted Mean Validity				.02								
K				1								
L				1								
M				1								
N Range				47								
<u>Rate Control</u>												
Median Validity					.02							
Weighted Mean Validity					.02							
K					1							
L					1							
M					1							
N Range					381							
<u>Finger Dexterity</u>												
Median Validity	.16			.19	.32			.20				
Weighted Mean Validity	.12			.19	.32			.17				
K	2			1	1			1				
L	18			1	1			8				
M	1			1	1			1				
N Range	50-234(71)			69	54			50-152(84)				
<u>Manual Dexterity</u>												
Median Validity	-.02				.36			.24				
Weighted Mean Validity	-.03				.36			.25				
K	2				1			2				
L	18				1			9				
M	1				1			1				
N Range	50-234(76)				54			50-212(86)				

(Continued)

Table 25 (continued)
Validity Summary for All Miscellaneous Jobs (N = 91 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	
<u>Wrist-Finger Speed</u>											
Median Validity	.16				.34			.30			
Weighted Mean Validity	.17				.34			.24			
K	2				1			2			
L	18				1			9			
M	1				1			1			
N Range	50-234(72)				54			50-212(86)			
<u>Aiming</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Arm-Hand Steadiness</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Speed of Arm Movement</u>											
Median Validity											
Weighted Mean Validity											
K											
L											
M											
N Range											
<u>Complex Psychomotor Predictors</u>											
Median Validity											
Weighted Mean Validity							.75				
K							.75				
L							1				
M							1				
N Range							1				

Table 26

Validity Summary for All Military Jobs (N = 547 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal		
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
PREDICTOR CONSTRUCT												
<u>Multilimb Coordination</u>												
.24			.22	.04	.26	.15	.15	.14				
Median Validity			.20	.04	.27	.13	.18	.15				
Weighted Mean Validity			3	1	14	1	1	1				
K			37	2	175	50	9	2				
L			5	2	14	5	1	2				
M												
N Range	730		29-1000(119)	381	29-7093(423)	32-574(311)	33-165(115)	27-32(30)				
<u>Control Precision</u>												
			.02		.19	.02						
Median Validity			.01		.20	.06						
Weighted Mean Validity			1		5	1						
K			2		27	3						
L			1		3	1						
M												
N Range			47-562(304)		74-3308(624)	331-459(403)						
<u>Rate Control</u>												
				.02	.12	.06						
Median Validity				.02	.11	.06						
Weighted Mean Validity			1	1	1	1						
K			1	1	6	15						
L			1	1	2	1						
M												
N Range				381	387-434(407)	164-311(311)						
<u>Finger Dexterity</u>												
			.14	-.12	.09	.04		-.32				
Median Validity			.11	-.12	.09	.05		-.32				
Weighted Mean Validity			2	1	5	1		1				
K			10	2	44	23		1				
L			2	1	3	2		1				
M								32				
N Range			51-1000(1000)	38	87-4779(547)	164-523(311)						
<u>Manual Dexterity</u>												
					.22							
Median Validity					.13							
Weighted Mean Validity					1							
K					2							
L					1							
M					1							
N Range					242-1496(469)							

(Continued)

Table 26 (continued)

Validity Summary for All Military Jobs (N = 547 Validity Coefficients)

PREDICTOR CONSTRUCT	CRITERION CONSTRUCT										
	Educational and School Achievement		Training Performance				Job Proficiency			Job Involvement/Withdrawal	
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination of Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction/Withdrawal
<u>Wrist-Finger Speed</u> Median Validity Weighted Mean Validity K L M N Range				.10 .10 1 1 1 1 79	-.19 -.19 1 1 1 1 62	.07 .07 2 2 1 1 244-245(246)					
<u>Aiming</u> Median Validity Weighted Mean Validity K L M N Range						.13 .13 3 7 5 244-728(246)					
<u>Arm-Hand Steadiness</u> Median Validity Weighted Mean Validity K L M N Range				.04 .06 1 6 1 1000		.09 .08 3 19 6 38-1942(183) 86-144(101)	.06 .06 1 3 1				
<u>Speed of Arm Movement</u> Median Validity Weighted Mean Validity K L M N Range						.10 .10 1 1 1 1194					
<u>Complex Psychomotor Predictors</u> Median Validity Weighted Mean Validity K L M N Range				-.12 -.12 1 1 1 54	.17 .17 1 1 1 48	.13 .09 3 47 4 63-293(139) 16-147(26)	.10 .13 3 46 5				

Table 27

Validity Summary for All Nonmilitary Jobs (N = 1826 Validity Coefficients)

	CRITERION CONSTRUCT											
	Educational and School Achievement		Training Performance					Job Proficiency				Job Involvement/ Withdrawal
	Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	
PREDICTOR CONSTRUCT												
Multilimb Coordination												
Median Validity				.06			-.05		-.16	.10		
Weighted Mean Validity				.06			-.00		-.16	.10		
K				1			1		1	1		
L				4			14		2	4		
M				2			2		1	1		
N Range				17			17		200	200		
Control Precision												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Rate Control												
Median Validity									-.02	.02		
Weighted Mean Validity									-.02	.02		
K									1	1		
L									2	4		
M									1	1		
N Range									301	301		
Finger Dexterity												
Median Validity	.13	.20	.28	.32	.22	.16	.12	.20	.19	.17		.14
Weighted Mean Validity	.12	.22	.28	.30	.19	.16	.18	.18	.16	.12		.14
K	3	1	1	1	2	1	3	4	1	3		1
L	96	25	1	8	2	1	6	406	8	28		1
M	4	1	1	1	1	1	3	3	1	3		1
N Range	10-271(62)	140-202(62)	75	46-86(63)	54-81(68)	80	49-430(100)	126-322(56)	41-130(58)	13-270(54)		98
Manual Dexterity												
Median Validity	.07	.21	.21	.26	.24	.20	.07	.22	.15	.18		.02
Weighted Mean Validity	.06	.20	.21	.28	.22	.20	.08	.20	.12	.15		.07
K	3	2	1	1	1	1	2	6	1	4		2
L	97	27	1	8	2	1	4	410	8	30		1
M	5	2	1	1	1	1	4	26	1	15		2
N Range	30-271(62)	140-202(62)	75	46-86(63)	54-81(68)	80	49-430(100)	126-322(56)	41-130(58)	13-270(54)		98-136(115)

(Continued)

Table 27 (continued)

Validity Summary for All Nonmilitary Jobs (N = 1826 Validity Coefficients)

CRITERION CONSTRUCT												
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/Withdrawal		
		Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
PREDICTOR CONSTRUCT												
Wrist-Finger Speed												
Median Validity		.11	.19	.32	.29	.07	.18	.20	.21			
Weighted Mean Validity		.12	.20	.32	.29	.05	.19	.20	.23			
K		2	1	1	1	1	3	1	1			
L		136	25	2	1	2	408	8	19			
M		3	1	1	1	1	1	1	1			
N Range		00-1079(62)	46-86(62)	54-81(68)	80	49-145(97)	30-322(56)	41-130(58)	32-100(52)			
Aiming												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Arm-Hand Steadiness												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Speed of Arm Movement												
Median Validity												
Weighted Mean Validity												
K												
L												
M												
N Range												
Complex Psychomotor Predictors												
Median Validity			.14	.21				.00	.02			-.09
Weighted Mean Validity			.18					-.03				
K			2				1	1	1	1	1	1
L			3				3	3	6	3	3	1
M			2				3	3	3	3	3	1
N Range								300-303(300)	300-303(300)	300-303(300)	63	63

infantrymen, mechanical or electronics repair specialists, or clerical or administrative specialists. Thus, these validity results may have little or no relevance for a number of large job families within the Army. Nevertheless, the many moderate to high validities in Table 26 demonstrate that psychomotor abilities have been valid predictors for the military in the past and should be carefully considered for inclusion in new selection and classification batteries.

Table 27 shows that virtually all nonmilitary psychomotor validity research has focused on determining the criterion-related validity of finger dexterity, manual dexterity, and wrist-finger speed. Of course, these are the three psychomotor abilities assessed by the GATB. Indeed, of the 1,826 validity coefficients tabulated for Table 27, 1,734 (95.0%) are from the GATB.

The median validities reported in Table 27 illustrate that finger dexterity, manual dexterity, and wrist-finger speed are valid predictors of a number of criteria. For example, all three abilities have demonstrated validity as predictors of various subjective criteria (e.g, instructor evaluations of educational and school achievement, subjective measures of training performance, and subjective measures of job proficiency), with most median validities for these criteria ranging from .20 to .30. Median correlations between these three abilities and job-related and archival measures of job proficiency are also quite high, averaging near .20. Median validities for school grades are somewhat lower, ranging from .07 to .13.

With the exception of multilimb coordination, very little information is available regarding the criterion-related validity of any other psychomotor abilities in nonmilitary research settings. Moreover, the validity summary for multilimb coordination is based on results from just two studies, one involving cab drivers and the other involving civilian pilots. Two different criterion measures were used in each study. Validities were quite low across all four criteria, ranging from -.16 to .10.

Comparison of Tables 26 and 27 is interesting primarily because it illustrates the differences in focus between military and nonmilitary psychomotor validity research. As noted previously, the military has invested most of its research effort in coordination and tracking abilities, while most nonmilitary psychomotor research has been concentrated on dexterity and manipulation abilities.

Since planes and pilots probably represent the military's greatest single investments in hardware and human resources, respectively, it is easy to understand why the military has directed most of its psychomotor validity research towards improving pilot and aircrew selection. Small improvements in these selection procedures can result in great savings in equipment and training costs. The job of pilot, however, is probably one of the few occupations requiring high levels of coordination and precise tracking and control adjustment skills.

In contrast, most nonmilitary research has been directed towards assessing the criterion-related validity of dexterity abilities, such as finger dexterity, manual dexterity, and wrist-finger speed. These are the psychomotor abilities which seem to be most relevant to clerical, indus-

trial, and maintenance and repair jobs. Typically, these jobs do not require the precise coordination of limb movements which the job of pilot requires. Nor do they require continuous adjustment to changing stimulus conditions. Instead, most of these jobs seem to demand moderately precise, moderately rapid, repetitive object manipulations. The specific manipulations required are often affected little by events in the environment, so there is no need to monitor and adjust motor responses in accordance with these environmental changes. Therefore, it is not surprising that nonmilitary research has focused on validating tests of these dexterity abilities.

The validity results presented in Tables 26 and 27 lend support to these observations. For example, median validities for multilimb coordination for military studies average .20, while those for nonmilitary studies average .00. Validity results are exactly opposite for finger dexterity. Median validities for nonmilitary studies average .20, while those for military studies average less than .10. These median validities point to differences in the psychomotor demands of the military (i.e., pilots) and nonmilitary (i.e., clerical, industrial, and maintenance and repair) jobs studied.

Validity Summary for All Jobs. Table 28 provides a summary of all of the validity coefficients identified during the literature review. The table contains very little new information, but it does provide a concise overview of the criterion-related validity results reported in the literature. Generally, these results indicate that psychomotor abilities are valid predictors of educational and school achievement, training performance, and job proficiency. Indeed, of the 10 abilities summarized in Table 28, only arm-hand steadiness fails to attain moderate validity (i.e., median $r = .10$) for at least one of the 12 criterion constructs. Additional analysis of this table reveals that, of the 33 predictor-criterion cells containing at least five validity observations (i.e., $L \geq 5$), median validities range from .20 to .29 in nine cells (27.3%), and from .10 to .19 in 18 cells (54.5%). Median validities are less than .10 in only six cells (18.2%).

Comparison of Criterion-Related Validities by Job Type and Criterion Category. Table 29 provides a summary of most of the validity evidence presented previously in Tables 8-28. The table was designed to aid in the comparison of validities across job types and criterion categories for each of the ten psychomotor ability constructs.

The validity data suggest that different psychomotor abilities may be useful for prediction for different job types. For example, tests of multilimb coordination and control precision have relatively high validity for predicting educational and training criteria for professional, technical, and managerial jobs (recall that most of these validities are based on studies of pilots), while tests of finger dexterity, manual dexterity, and wrist-finger speed appear to be most useful for predicting educational and job proficiency criteria for industrial jobs.

The validity data also indicate very little difference in the predictability of different criterion categories. While the overall median validity for job proficiency criteria (median $r = .20$, $N = 1,358$ validity coefficients) is somewhat greater than the median validity for educational (me-

Table 28

Validity Summary for All Jobs (N = 2373 Validity Coefficients)

CRITERION: CONSTRUCT													
Educational and School Achievement		Training Performance					Job Proficiency				Job Involvement/ Withdrawal		
		Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures				
Course Grades		Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Co-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal	
PREDICTOR CONSTRUCT													
Multilimb Coordination													
Median Validity	.24			.21	.04	.25	.14	.15	-.06	.10			
Weighted Mean Validity	.24			.19	.04	.27	.12	.18	-.14	.10			
K	1			4	1	14	2	1	2	1			
L	1			41	2	175	64	9	4	4			
M	1			6	2	14	6	1	2	1			
N Range	730			29-1000(58)	381	29-709(34)	17-144(87)	33-165(115)	27-299(866)	299			
Control Precision													
Median Validity				.02		.19	.02						
Weighted Mean Validity				.01		.20	.06						
K				1		5	1						
L				2		27	3						
M				1		3	1						
N Range				47-562(304)		74-3308(614)	331-459(403)						
Rate Control													
Median Validity				.02		.12	.06		-.02	.02			
Weighted Mean Validity				.02		.11	.06		-.02	.02			
K				1		1	1		1	1			
L				1		6	15		2	4			
M				1		2	1		1	1			
N Range				381		387-434(607)	164-311(311)		301	301			
Finger Dexterity													
Median Validity	.13	.20	.28	.15	.00	.10	.04	.20	.17	.17		.14	
Weighted Mean Validity	.12	.22	.28	.13	.08	.09	.07	.18	.13	.12		.14	
K	3	1	1	3	2	6	4	4	2	3		1	
L	96	25	7	18	4	45	29	406	9	28		1	
M	4	1	1	3	2	4	5	3	2	3		1	
N Range	30-271(61)	40-202(62)	75	46-1000(82)	38-81(46)	80-4779(53)	49-523(311)	26-322(56)	32-130(57)	15-270(158)		98	
Manual Dexterity													
Median Validity	.07	.21	.21	.26	.24	.20	.07	.22	.15	.18		.02	
Weighted Mean Validity	.06	.20	.21	.28	.22	.13	.08	.20	.12	.15		.07	
K	3	2	1	1	1	2	2	6	1	4		2	
L	97	27	1	8	2	3	4	410	8	30		3	
M	5	2	1	1	2	3	4	4	1	4		2	
N Range	30-271(61)	40-202(63)	75	46-1000(82)	34-81(46)	80-1296(242)	49-145(76)	26-322(56)	41-130(58)	15-270(158)		98-134(134)	

(Continued)

Table 28 (continued)

Validity Summary for All Jobs (N = 2373 Validity Coefficients)

CRITERION CONSTRUCT											
Educational and School Achievement		Training Performance					Job Proficiency			Job Involvement/ Withdrawal	
Course Grades	Instructor Evaluations	Objective Measures	Subjective Measures	Combination Objective and Subjective Measures	Go-No Go Training Courses	Hands-On Measures	Subjective Measures	Job-Related Measures	Archival Measures	Job Satisfaction	Job Withdrawal
PREDICTOR CONSTRUCT											
Wrist-Finger Speed											
Median Validity	.19	.16	.30		.09	.07	.18	.20	.21		
Weighted Mean Validity	.20	.18	.16		.10	.05	.19	.20	.23		
K	1	2	2		3	1	3	1	1		
L	25	8	3		3	2	408	8	19		
M	1	2	2		2	1	1	1	1		
N Range	30-1079(69)	40-202(62)	46-86(63)	54-81(62)	80-245(244)	49-145(97)	30-122(56)	1-130(58)	32-100(52)		
Aiming											
Median Validity					.13						
Weighted Mean Validity					.13						
K					3						
L					7						
M					5						
N Range					244-728(244)						
Arm-Hand Steadiness											
Median Validity					.09	.06					
Weighted Mean Validity					.08	.06					
K		.06	.06		3	1					
L		1	6		19	3					
M		1	1		6	1					
N Range		1000			38-1942(185)	86-144(101)					
Speed of Arm Movement											
Median Validity					.10						
Weighted Mean Validity					.10						
K					1						
L					1						
M					1						
N Range					1194						
Complex Psychomotor Predictors											
Median Validity					.15	.11		.00	.02		-.09
Weighted Mean Validity					.09	.14		-.01	.03		-.09
K		.09	.17		3	5		1	1		1
L		2	1		47	49		3	6		1
M		3	1		4	7		3	1		1
N Range		54-107(107)	48		23-293(134)	6-127(26)		800-303(300)	300-300(300)		63

Table 29

Median Validities by Job Type and Criterion Category

Psychomotor Abilities and Criterion Categories	Job Type									
	Professional and Managerial	Clerical	Sales	Protective Services	Service	Mechanical and Structural Maintenance	Electronics	Industrial	Miscellaneous	All Job Types
<u>Multilimb Coordination</u>										
Education	.24 (1)	--	--	--	--	--	--	--	.18 (4)	.24 (1)
Training	.20 (240)	--	--	.20 (38)	--	--	--	--	.20 (282)	.20 (282)
Job Proficiency	--	--	--	.15 (11)	.06 (6)	--	--	--	--	.14 (17)
Job Involvement	--	--	--	--	--	--	--	--	--	--
Overall	.21 (241)	--	--	.20 (49)	.06 (6)	--	--	--	.18 (4)	.20 (300)
<u>Control Precision</u>										
Education	--	--	--	--	--	--	--	--	--	--
Training	.19 (24)	--	--	.14 (7)	--	--	--	--	.02 (1)	.17 (32)
Job Proficiency	--	--	--	--	--	--	--	--	--	--
Job Involvement	--	--	--	--	--	--	--	--	--	--
Overall	.19 (24)	--	--	.14 (7)	--	--	--	--	.02 (1)	.17 (32)
<u>Rate Control</u>										
Education	--	--	--	--	--	--	--	--	--	--
Training	.08 (17)	--	--	.05 (4)	--	--	--	--	.02 (1)	.06 (22)
Job Proficiency	--	--	--	--	-.02 (6)	--	--	--	--	-.02 (6)
Job Involvement	--	--	--	--	--	--	--	--	--	--
Overall	.08 (17)	--	--	.05 (4)	-.02 (6)	--	--	--	.02 (1)	.06 (28)
<u>Finger Dexterity</u>										
Education	.11 (57)	.27 (2)	--	--	.22 (12)	.18 (18)	.17 (3)	.28 (11)	.16 (19)	.16 (121)
Training	.07 (59)	.08 (1)	--	.09 (16)	.08 (6)	.25 (6)	--	.11 (7)	.26 (2)	.08 (97)
Job Proficiency	.13 (47)	.12 (44)	.08 (5)	.07 (6)	.13 (33)	.21 (39)	.15 (7)	.24 (254)	.20 (8)	.19 (443)
Job Involvement	--	--	--	--	--	--	--	.14 (1)	--	.14 (1)
Overall	.10 (163)	.13 (47)	.08 (5)	.09 (22)	.14 (51)	.20 (63)	.16 (10)	.24 (273)	.18 (28)	.16 (662)

(Continued)

Table 29 (continued)

Median Validities by Job Type and Criterion Category

Psychomotor Abilities and Criterion Categories	Job Type									
	Professional, Technical, and Managerial	Clerical	Sales	Protective Services	Service	Mechanical and Structural Maintenance	Electronics	Industrial	Miscellaneous	All Job Types
<u>Manual Dexterity</u>										
Education	.07 (57)	.32 (2)	--	--	.17 (12)	.20 (18)	.04 (3)	.26 (14)	-.02 (18)	.10 (124)
Training	.20 (3)	.13 (1)	--	--	.29 (2)	.20 (4)	--	.17 (7)	.36 (1)	.19 (18)
Job Proficiency	.16 (47)	.12 (44)	.08 (5)	.15 (5)	.18 (33)	.21 (40)	.12 (7)	.24 (258)	.24 (9)	.22 (448)
Job Involvement	---	--	--	--	--	--	--	.02 (3)	--	.02 (3)
Overall	.12 (107)	.13 (47)	.08 (5)	.15 (5)	.18 (47)	.20 (62)	.08 (10)	.24 (282)	.10 (28)	.19 (593)
<u>Wrist-Finger Speed</u>										
Education	.10 (100)	.32 (2)	--	--	.22 (12)	.18 (18)	.12 (3)	.12 (8)	.16 (18)	.13 (161)
Training	.09 (3)	.25 (1)	--	--	.06 (4)	.06 (3)	--	.16 (4)	.34 (1)	.12 (16)
Job Proficiency	.14 (48)	.14 (46)	.14 (5)	.23 (5)	.20 (33)	.16 (38)	.16 (7)	.22 (244)	.30 (9)	.18 (435)
Job Involvement	--	--	--	--	--	--	--	--	--	--
Overall	.11 (151)	.16 (49)	.14 (5)	.23 (5)	.20 (49)	.16 (59)	.14 (10)	.21 (256)	.22 (28)	.17 (612)
<u>Aiming</u>										
Education	--	--	--	--	--	--	--	--	--	--
Training	.13 (7)	--	--	--	--	--	--	--	--	.13 (7)
Job Proficiency	--	--	--	--	--	--	--	--	--	--
Job Involvement	--	--	--	--	--	--	--	--	--	--
Overall	.13 (7)	--	--	--	--	--	--	--	--	.13 (7)
<u>Arm-Hand Steadiness</u>										
Education	--	--	--	--	--	--	--	--	--	--
Training	.07 (25)	--	--	.06 (3)	--	--	--	--	--	.06 (28)
Job Proficiency	--	--	--	--	--	--	--	--	--	--
Job Involvement	--	--	--	--	--	--	--	--	--	--
Overall	.07 (25)	--	--	.06 (3)	--	--	--	--	--	.06 (28)

(Continued)

Table 29 (continued)

Median Validities by Job Type and Criterion Category

Psychomotor Abilities and Criterion Categories	Job Type									
	Professional, Technical and Managerial	Clerical	Sales	Protective Services	Service	Mechanical and Structural Maintenance	Electronics	Industrial	Miscellaneous	All Job Types
<u>Speed of Arm Movement</u>										
Education	—	—	—	—	—	—	—	—	—	—
Training	.10 (1)	—	—	—	—	—	—	—	—	.10 (1)
Job Proficiency	—	—	—	—	—	—	—	—	—	—
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.10 (1)	—	—	—	—	—	—	—	—	.10 (1)
<u>Complex Psychomotor Predictors</u>										
Education	—	—	—	—	—	—	—	—	—	—
Training	.16 (57)	—	—	.07 (38)	.02 (2)	—	—	.14 (2)	.75 (1)	.13 (100)
Job Proficiency	—	—	—	—	.00 (9)	—	—	—	—	.00 (9)
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.16 (57)	—	—	.07 (38)	.00 (11)	—	—	.06 (3)	.75 (1)	-.09 (1)
										.10 (110)
<u>All Psychomotor Abilities</u>										
Education	.10 (215)	.32 (6)	—	—	.21 (36)	.18 (54)	.10 (9)	.26 (33)	.11 (54)	.13 (407)
Training	.16 (436)	.13 (3)	—	.12 (106)	.09 (14)	.21 (13)	—	.15 (20)	.30 (11)	.15 (603)
Job Proficiency	.14 (142)	.13 (134)	.08 (15)	.15 (27)	.13 (120)	.20 (117)	.15 (21)	.23 (756)	.26 (26)	.20 (1358)
Job Involvement	—	—	—	—	—	—	—	.02 (5)	—	.02 (5)
Overall	.14 (793)	.14 (143)	.08 (15)	.13 (133)	.14 (170)	.19 (184)	.13 (30)	.23 (814)	.18 (91)	.17 (2373)

NOTE: Each entry in this table consists of two numbers. The decimal number is the median validity coefficient for that particular psychomotor ability - criterion category - job type combination. The number in parentheses which follows is the number of validity coefficients tabulated in computing the median.

dian $r=.13$, $N=407$ validity coefficients) and training (median $r=.15$, $N=603$ validity coefficients) criteria, the results vary considerably from job type to job type and from ability to ability.

Table 29 shows that most of the psychomotor abilities have been used successfully as predictors (i.e., median $r>.20$) in at least one job type. For example, multilimb coordination has been an effective predictor for professional, technical, and managerial jobs and for protective service jobs. Manual dexterity has demonstrated validity for service jobs and industrial jobs. For those abilities which have not evidenced their value as predictors, the problem may be a lack of research. For example, in total only 28 validity coefficients are available for rate control, only seven are available for aiming, only 28 are available for arm-hand steadiness, and only one is available for speed of arm movement. Additional research may show that tests of these abilities are also effective predictors for some job types.

Finally, Table 29 suggests that psychomotor abilities may be useful predictors only for certain job types. For example, across all abilities and criteria, the median validity for psychomotor abilities for industrial jobs is .23 ($N=814$ validity coefficients). Psychomotor abilities have also demonstrated moderate validity for professional, technical, and managerial jobs (median $r=.14$, $N=793$ validity coefficients), clerical jobs (median $r=.14$, $N=143$ validity coefficients), protective service jobs (median $r=.13$, $N=133$ validity coefficients), service jobs (median $r=.14$, $N=170$ validity coefficients), mechanical and structural maintenance jobs (median $r=.19$, $N=184$ validity coefficients), electronics jobs (median $r=.13$, $N=30$ validity coefficients), and miscellaneous jobs (median $r=.18$, $N=91$ validity coefficients). These job types would include a number of Army MOS. On the other hand, the median validity for sales jobs was only .08 ($N=15$ validity coefficients), indicating that psychomotor ability tests would have no utility as predictors for this nonmilitary job type.

Summary of the Validity Evidence

Past research demonstrates that psychomotor ability tests can be valid predictors of educational, training, and job proficiency criteria for many job types. Indeed, the validity data presented in Tables 7-29 suggest that sales may be one of the only job types for which no psychomotor abilities have any predictive validity.

Test development efforts should be guided by a content-oriented test development strategy. Military psychologists who have adopted this strategy in the past (e.g., Eggenberger, 1976; Guilford & Lacey, 1947; Melton, 1947) have met with good success in predicting job and training performance. Moreover, content valid tests typically have a great deal of face validity, increasing the likelihood that the tests will be accepted by users and test subjects.

In sum, the major questions regarding the usefulness of psychomotor tests for personnel selection and classification have little to do with the tests' validity (cf. North & Griffin, 1977; Passey & McLaurin, 1966; Zeidner, Martinek, and Anderson, 1961). Rather, the questions concern special psychomotor testing issues (e.g., test reliability, cost effectiveness/utility of psychomotor testing, group differences in psycho-

motor test scores, the fidelity between psychomotor tests and the criteria they are intended to predict). These questions are addressed in the next section.

SECTION V

CURRENT ISSUES IN RESEARCH ON PSYCHOMOTOR ABILITIES

Stability and Differential Stability

One of the major areas of current psychomotor ability research is the stability of individual differences in psychomotor test scores as the test task is practiced.

Some of the issues involved in this research have been noted previously in the discussion of Fleishman's distinction between skills and abilities. Recall that Fleishman (1960; Fleishman & Hempel, 1954b, 1955; Parker & Fleishman, 1962) and others (e.g., Adams, 1953, 1957; Hinrichs, 1970) have found that the abilities accounting for variance in skilled motor performance change with practice. In general, these researchers have found that correlations between cognitive abilities (e.g., verbal ability, spatial relations) and task proficiency decrease as the task is practiced, while correlations between psychomotor (e.g., speed of arm movement) and basic perceptual abilities (e.g., reaction time, response orientation) and task proficiency increase slightly. In addition, as the task is practiced, there is an ever increasing amount of variance specific to the task itself (Fleishman, 1975).

This skills-abilities research suggests that the methods typically used to assess psychomotor abilities for criterion-related validity research may be deficient. In most such research, subjects are given at most one or two practice trials before the tests are administered (e.g., Melton, 1947; U.S. Department of Labor, 1970). Typically, tests then span no more than four to six trials and require only about 5 minutes of actual testing time. Fleishman's research indicates that the resulting test scores are likely to be at least somewhat contaminated with variance attributable to cognitive abilities. Most of this variance could be eliminated if subjects were permitted more practice time prior to test administration.

Over the past several years, the Naval Medical Research and Development Command has sponsored a number of studies as part of an effort to study the effects of adverse environments (e.g., vibration and ship motion) on task performance (Kennedy & Bittner, 1977; Kennedy, Carter, & Bittner, 1980). Research on the impact of environmental effects almost always requires the use of repeated-measures designs to assess pre-, per-, and post-exposure effects on task performance. Results from repeated-measures designs can be misleading, however, if performance is measured on a task that shows practice effects. Therefore, an initial goal of this research has been to identify a battery of Performance Evaluation Tests for Environmental Research (PETER) which can be used in repeated-measures studies of different environments.

According to Jones, Kennedy, and Bittner (1981), tests which are included in the PETER battery must stabilize over time. That is, the test task may show practice effects initially, but "there (must) come a point...after which practice effects no longer appear" (p. 143). Bittner (1979) has called this criterion differential stability. "Differential stability refers to the point in practice after which individual learning curves are parallel and at most slowly increasing, except for random error;

after this point all intertrial (interday) correlations are the same except for random sampling variations" (Kennedy, Bittner, & Jones, 1981, p. 310). Differential stability also implies that the variance among subjects' test scores remains constant across trials and that the correlations between any given early, "unstabilized" trial and all stabilized trials are approximately equal (Jones, Kennedy, & Bittner, 1981).

Since not all tasks stabilize (i.e., show differential stability), the PETER researchers were first interested in identifying a stable set of tasks for the PETER battery. In addition, for tasks that were stable, PETER researchers sought to determine how much practice was required before differential stability was attained.

To date, PETER differential stability research has focused on broad cognitive abilities, such as grammatical reasoning (Carter, Kennedy, & Bittner, 1981) and arithmetic computation (Seales, Kennedy, & Bittner, 1980), and complex perceptual-psychomotor tasks modeled in many instances after actual Navy job duties. These include a navigation plotting task (Wiker, Kennedy, & Pepper, 1983), a number of Atari video games (Jones, Kennedy, & Bittner, 1981; Kennedy, Bittner, Harbeson, & Jones, 1982), and a compensatory tracking task (Kennedy, Bittner, & Jones, 1981).

Almost all of the tasks the PETER researchers have examined have eventually demonstrated differential stability. In most cases, however, considerable practice was required before the tasks stabilized. For example, the Atari Air Combat Maneuvering (ACM) game did not attain differential stability until after subjects had practiced 30 minutes per day for six days--and the ACM game became stable more quickly than any of the other perceptual-psychomotor tasks examined.

These results may be surprising in light of the high psychomotor test reliabilities reported by Melton (1947), Fleishman (1958), and others. Reliabilities for some of the more commonly used psychomotor tests are reported in Table 30. Most test-retest reliabilities for intertest durations of one month or less are between .70 and .90, indicating fairly good reliability for these psychomotor tests. How can these results be reconciled with the differential stability results obtained by the PETER researchers?

First, it is important to note that differential stability is not the same as high intertrial correlations or high test-retest reliability. Indeed, the average correlation among stabilized trials could be .60, .40, or even .20 when differential stability is attained. Differential stability does not imply that measures are relatively free of random, unreliable variance. Rather, it means that the percentage of task performance variance which is reliable (i.e., common across days or trials) does not change from day to day or trial to trial once the task has stabilized.

Second, test-retest reliabilities reported by the PETER researchers are not appreciably different from those reported by Melton (1947), Fleishman (1958), and others. Reliabilities for some of the PETER psychomotor tests are summarized in Table 31. The median reliability for stabilized trials for these seven tests is .80. This is very similar to the median test-retest reliability of .78 for the psychomotor tests summarized in Table 30. (This median is based on only those reliabilities reported

Table 30

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
Arm-Mand Steadiness Test	.76	Mean inter-trial correlation for 3 preferred-hand trials, corrected for the full length of the test	N=310 unclassified candidates for pilot training	Melton, 1947
	.75	1-week test-retest reliability	N=328 unclassified candidates for pilot training	Melton, 1947
Complex Coordination Test (Melton, 1947)	.89	Odd-even split-half reliability, corrected for the full length of the test (4 trials)	N=7627 unclassified candidates for pilot training	Melton, 1947
	.82	Odd-even split-half reliability, corrected for the full length of the test (2 trials)	N=204 basic trainee airmens	Fleishman, 1958
	.87	Immediate test-retest reliability	N=415 unclassified candidates for pilot training	Melton, 1947
	.59	1-week test-retest reliability	N=313 unclassified candidates for pilot training	Melton, 1947
	.83	28-day test-retest reliability	N=692 unclassified candidates for pilot training	Melton, 1947
Complex Coordination Test (Sanders et al., 1971)	.92	Correlation between the Generated Score during minutes 4 and 5 of the test, corrected for length to represent the reliability of the sum of the two scores	N=120 Air Force officer trainees	Sanders et al., 1971

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
GATB Finger Dexterity Scale	.65	3-month test-retest reliability	N=605 male high school seniors	U. S. Department of Labor, 1970
	.69	3-month test-retest reliability	N=605 female high school seniors	U. S. Department of Labor, 1970
	.76	1-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
	.68	2-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
	.74	3-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
GATB Manual Dexterity Scale	.73	3-month test-retest reliability	N=605 male high school seniors	U. S. Department of Labor, 1970
	.72	3-month test-retest reliability	N=605 female high school seniors	U. S. Department of Labor, 1970
	.76	1-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
	.72	2-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
	.78	3-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
GATB Motor Coordination Scale	.76	3-month test-retest reliability	N=605 male high school seniors	U. S. Department of Labor, 1970
	.86	3-month test-retest reliability	N=605 female high school seniors	U. S. Department of Labor, 1970
	.86	1-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
	.85	2-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
	.88	3-year test-retest reliability	Government employees (N unknown)	U. S. Department of Labor, 1970
Minnesota Rate of Manipulation	.87	Reliability of the placing score, method unknown; since the test includes two placing trials, it is likely that this reliability represents the correlation between these two trials, corrected for the total length of the test	N=203 freshman and sophomore Air Force ROTC students	Parker and Fleishman, 1960
	.79	Reliability of the turning score, method unknown; since the test includes two placing trials, it is likely that this reliability represents the correlation between these two trials, corrected for the total length of the test	N=203 freshman and sophomore Air Force ROTC students	Parker and Fleishman, 1960

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
Motor Judgment Test	.96	Odd-even split-half reliability for the errors score, corrected for the full length of the test (8 trials)	N=50 qualified aircrew candidates	Melton, 1947
	.92	Odd-even split-half reliability for the revolutions score, corrected for the full length of the test (8 trials)	N=50 qualified aircrew candidates	Melton, 1947
	.76	Odd-even split-half reliability for the ratio of errors to revolutions, corrected for the full length of the test (4 trials)	N=204 basic trainee airmen	Fleishman, 1958
O'Connor Finger Dexterity Test	.76	Unknown	N=203 freshman and sophomore Air Force ROTC students	Parker and Fleishman, 1960
Purdue Pegboard Test	.90	Average reliability of the four subparts of the test, corrected for a fourfold increase in length; method of estimating the reliability of the four subparts is unknown	N=203 freshman and sophomore Air Force ROTC students	Parker and Fleishman, 1960
Rate Control Test	.81	Odd-even split-half reliability, corrected for the full length of the test (8 trials)	N=381 airborne radar students	Melton, 1947
	.69	Odd-even split-half reliability, corrected for the full length of the test (4 trials)	N=204 basic trainee airmen	Fleishman, 1958

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
Rotary Pursuit Test	.98	Odd-even split-half reliability, corrected for the full length of the test (20 trials)	N=301 unclassified candidates for pilot training	Melton, 1947
	.81	Odd-even split-half reliability, corrected for the full length of the test (5 trials)	N=204 basic trainee airmen	Fleishman, 1958
	.88	Immediate test-retest reliability	N=398 unclassified candidates for pilot training	Melton, 1947
Rudder Control Test	.93	Odd-even split-half reliability, corrected for the full length of the test (6 trials)	N=1000 unclassified candidates for pilot training	Melton, 1947
	.82	Odd-even split-half reliability, corrected for the full length of the test (3 trials)	N=204 basic trainee airmen	Fleishman, 1958
	.67	28-day test-retest reliability	N=311 unclassified candidates for pilot training	Melton, 1947
	.76	28-day test-retest reliability	N=312 unclassified candidates for pilot training	Melton, 1947

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
Santa Ana Finger Dexterity Test	.93	Odd-even split-half reliability, corrected for the full length of the test (5 trials)	N=1000 unclassified candidates for pilot training	Melton, 1947
	.85	Immediate test-retest reliability	N=403 unclassified candidates for pilot training	Melton, 1947
	.74	1-week test-retest reliability	N=314 unclassified candidates for pilot training	Melton, 1947
	.74	28-day test-retest reliability	N=701 unclassified candidates for pilot training	Melton, 1947
Single-Dimension Pursuitmeter	.88	Odd-even split-half reliability for the timer score, corrected for the full length of the test (8 trials)	N=1483 unclassified candidates for pilot training	Melton, 1947
	.92	Odd-even split-half reliability for the work adder score, corrected for the full length of the test (8 trials)	N=1483 unclassified candidates for pilot training	Melton, 1947
	.76	Odd-even split-half reliability for the timer score, corrected for the full length of the test (4 trials)	N=204 basic trainee airmen	Fleishman, 1958

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
Steadiness Aiming Test	.96	Average inter-trial correlation between scores on the six trials of the test, corrected for the full length of the test	N=461 aviation students	Melton, 1947
	.84	Unknown	N=203 freshmen and sophomore Air Force ROTC students	Parker and Fleishman, 1960
Two-Hand Coordination Test (Melton, 1947)	.90	Odd-even split-half reliability, corrected for the full length of the test (8 trials)	N=1912 unclassified candidates for pilot training	Melton, 1947
	.80	Odd-even split-half reliability, corrected for the full length of the test (4 trials)	N=204 basic trainee airmen	Fleishman, 1958
	.83	Immediate test-retest reliability	N=416 unclassified candidates for pilot training	Melton, 1947
	.78	1-week test-retest reliability	N=320 unclassified candidates for pilot training	Melton, 1947
	.87	28-day test-retest reliability	N=700 unclassified candidates for pilot training	Melton, 1947

(Continued)

Table 30 (Continued)

Reliabilities for Selected Psychomotor Tests

Test	Reliability	Method of Assessment	Sample	Reference
Two-Hand Coordination Test (Sanders et al., 1971)	.81	Correlation between the General Score during minutes 4 and 5 of the test, corrected for length to represent the reliability of the sum of the two scores	N=120 Air Force officer trainees	Sanders et al., 1971
Two-Hand Pursuit Test	.90	Mean intertrial correlation, corrected for the full length of the test (8 trials)	N=2112 unclassified candidates for pilot training	Melton, 1947
	.83	Odd-even split-half reliability, corrected for the full length of the test (4 trials)	N=204 basic trainee airmen	Fleishman, 1958
Two-Plate Tapping Test	.96	Mean correlation between test scores for each of the first three minutes of the test, corrected for the full length of the test (8 minutes)	N=500 unclassified candidates for pilot training	Melton, 1947
	.99	Unknown	N=203 freshmen and sophomore Air Force ROTC students	Parker and Fleishman, 1960

Table 31

Reliability^a of PETER Psychomotor Tests

Test	Reliability	Sample Size ^b	Reference
Atari Air Combat Maneuvering	.93	N=22	Jones, Kennedy, and Bittner, 1981
Atari Slalom	.60	N=15	Kennedy and Bittner, 1980
Compensatory Tracking	.78	N=18	Kennedy, Bittner, and Jones, 1981
Critical Tracking	.85	N=18	Damos, Kennedy, and Bittner, 1980
Spoke Task	.80	N=18	Bittner, Lundy, Kennedy, and Harbeson, 1982
Tapping	.85	N=18	Kennedy, Bittner, and Einbender, 1980
Trail Making	.40	N=18	Kennedy, Bittner, and Einbender, 1980

a. The reliability represents the average correlation between stabilized daily test scores. This reliability is actually a form of test-retest reliability, with the interval between testing sessions ranging from 1 to 14 days, depending on the particular test and the number of days required for the test to stabilize.

b. All subjects were Navy enlisted men.

for intertest durations of one month or less.)

Nevertheless, the PETER researchers have suggested that task performance be assessed only after the task has stabilized. In part, this is because data gathered via repeated-measures designs are much more difficult to analyze if differential stability has not been attained (Jones, Kennedy, & Bittner, 1981). In part, however, it also reflects a concern that the task is changing and/or the abilities used by the subject to perform the task are changing prior to the time at which the task stabilizes (cf. Hulin & Alvares, 1971). If either or both of these is true, then the task may not be measuring the same ability or abilities across assessment trials.

While this may at first glance appear to be a serious problem which jeopardizes the criterion-related validity of many psychomotor tests, evidence from a number of sources suggests that this really is not a threat to the use of psychomotor tests as predictors of job performance.

First, even without differential stability, test-retest reliability for the PETER tests remains quite high across 15 days (e.g., .78 for the Navigation Plotting task and .69 for the Atari Air Combat Maneuvering task). These correlations suggest that the abilities required to perform an unstabilized task are by and large quite similar to the abilities required to perform the task after it has stabilized.

Second, Fleishman's research cited above indicates that the variance not shared by the unstabilized task and the stabilized task is primarily attributable to broad cognitive abilities such as verbal ability (Fleishman, 1975). To the extent that the criterion job performance measure is uncorrelated with cognitive abilities, traditional psychomotor ability assessment procedures may yield an attenuated estimate of the true correlation between psychomotor abilities and job performance criteria. Nevertheless, the test-retest reliabilities reported in Tables 30 and 31 and the research on ability correlates of psychomotor task performance cited previously indicate that performance on even the first few trials of a psychomotor task is heavily influenced by psychomotor abilities. For this reason, AAF researchers found that their psychomotor tests contributed unique variance to the prediction of pilot training attrition (Melton, 1947).

Thus, while stabilized tasks might be more valid as predictors of job performance for jobs which are highly dependent upon psychomotor abilities (i.e., because there will be relatively little variance in the predictor which is attributable to non-psychomotor abilities), unstabilized tasks will generally be almost as valid. In addition, assessment of unstabilized tasks is much more economical than assessment of stabilized tasks. Assessment of unstabilized tasks precludes the need to expend valuable subject time on task practice. If subject time is limited--as it almost always is--use of unstabilized tasks allows assessment of a much broader range of tasks and abilities.

Utility of Psychomotor Predictors

Lack of validity data for psychomotor tests and lack of information about the standard deviation of job performance in dollars for various Army job groups preclude a formal analysis of the utility of psychomotor tests

as guides for job classification of Army enlisted personnel (Brogden, 1946, 1949; Cronbach & Gleser, 1965). The validity information summarized previously, however, suggests that psychomotor tests will indeed be effective as predictors of job performance for many Army jobs. In addition, critical incident data collected for first-tour soldiers illustrates that ineffective job performance can be quite costly to the Army in terms of deadlined or irreparably damaged equipment and vehicles and personnel time lost to maintenance and repairs.

Nevertheless, there are at least two reasons that psychomotor testing might lack utility for the Army. First, it may be that psychomotor tests contribute little or no unique variance to the prediction of job performance. And, second, it may be that psychomotor testing--or, at the very least, psychomotor apparatus testing--is so expensive that any increments in the prediction of job performance are more than offset by increases in costs associated with testing. Each of these possibilities will be addressed below.

Correlations between Psychomotor and Cognitive-Perceptual Abilities. To the extent that psychomotor abilities are only moderately or weakly correlated with cognitive-perceptual abilities, the utility of psychomotor testing will be enhanced. This is because low correlations between psychomotor and cognitive-perceptual abilities indicate that psychomotor abilities can contribute unique variance to the prediction of job and training performance. If, however, the correlations between psychomotor abilities and cognitive-perceptual abilities are high, it will mean that psychomotor ability tests would contribute very little above and beyond cognitive ability tests to the prediction of job and training performance. Since psychomotor testing tends to be more expensive than cognitive-perceptual testing (i.e., because of the need for special test apparatus and/or individual test administration), high correlations between psychomotor and cognitive-perceptual abilities would suggest that psychomotor testing would not be worthwhile to the Army. Under those conditions, the payoff from psychomotor testing in terms of improved prediction and classification would be negligible at best.

To determine typical correlations between various psychomotor and cognitive-perceptual abilities, the articles, reports, and manuals identified during the literature search were examined in an effort to locate intercorrelation matrixes that included both psychomotor and cognitive-perceptual tests. Five such sources were identified (Fleishman, 1954b; Melton, 1947; Michael, 1949; Parker & Fleishman, 1960; U.S. Department of Labor, 1970). In total, these five sources contained 17 intercorrelation matrixes.

Each psychomotor test used in these studies was classified into one of the nine psychomotor ability constructs (complex psychomotor predictors were not included in this classification scheme) and each cognitive-perceptual ability test was classified into one of the 12 cognitive-perceptual ability constructs used in the validity summary tables in the literature review for cognitive abilities. In some cases, tests did not fit neatly into any of the 21 ability constructs and therefore were excluded from further analysis.

Correlations were then tabulated between each psychomotor ability and each cognitive-perceptual ability. In total, 1,066 correlations were tabu-

lated. Among the psychomotor abilities, the number of correlations ranged from 337 for multilimb coordination to zero for aiming. Within the cognitive-perceptual ability domain, the number of correlations ranged from 324 for spatial ability to zero for fluency.

Median correlations were then computed between each psychomotor ability and each cognitive-perceptual ability. Table 32 summarizes these median correlations. In general, the table shows that the two ability domains do not overlap greatly. Of the 61 median correlations reported in the table, only 10 are greater than .20, and over 40% (25) are .10 or less. Only four psychomotor abilities appear to be even moderately correlated with two or more cognitive-perceptual abilities.

Multilimb coordination is one of these four psychomotor abilities. Median correlations for multilimb coordination include .29 with mechanical aptitude ($N=45$ correlations) and .23 with perceptual speed and accuracy ($N=64$ correlations).

The moderately high correlation between multilimb coordination and mechanical aptitude may reflect a co-occurrence of learning experiences. That is, people who spend a lot of time building and/or repairing mechanical equipment are likely to learn mechanical principles. In addition, they are likely to spend a lot of time working in situations which require the careful coordination of two hands or hands and feet. Thus, the correlation between these two abilities is not surprising. Nor is the moderate correlation between multilimb coordination and spatial ability (median $r=.18$, $N=107$ correlations). Most tests of multilimb coordination require the subject to visualize and/or monitor a stimulus pattern. The moderate correlations of multilimb coordination with perceptual speed and accuracy and perception (median $r=.19$, $N=5$ correlations) may also be due to this perceptual requirement of multilimb coordination tests.

None of these median correlations is very high, however; all are less than .30. This suggests that there is still a great deal of variance in measures of multilimb coordination which is not tapped by cognitive-perceptual tests. Moreover, the validity summary tables described previously (e.g., Table 29) show that multilimb coordination has been a valid predictor of training performance for at least two job types. Therefore, the correlations between multilimb coordination and these three cognitive-perceptual ability do not pose a serious threat to the utility of testing for multilimb coordination.

Finger dexterity is the second psychomotor ability which is moderately correlated with at least two cognitive-perceptual abilities (median $r=.26$ with perception, $N=8$ correlations; median $r=.20$ with perceptual speed and accuracy, $N=29$ correlations). The moderate correlations between finger dexterity and these two cognitive-perceptual abilities are probably due to the perceptual and visualization demands of finger dexterity tasks. Further evidence for this is supplied by the median correlation of .16 between finger dexterity and spatial ability measures ($N=52$ correlations). Nevertheless, these median correlations are only moderate at best, and the median validity coefficients reported in Table 29 suggest that finger dexterity would be a valid predictor of educational and school achievement, training performance, and job proficiency in several job types. Thus, it would seem prudent to continue considering finger dexterity for inclusion

Table 32

Correlations Between Psychomotor and Cognitive-Perceptual Abilities

PSYCHOMOTOR ABILITIES	COGNITIVE-PERCEPTUAL ABILITIES										
	Spatial Ability	Perceptual Speed & Accuracy	Verbal Comprehension	Reasoning	Number Facility	Memory	Perception	Mechanical Aptitude	Automotive/Shop/ Tool Knowledge	Electronics Information	Science Knowledge
Multilimb Coordination	.18 (107)	.23 (64)	.08 (30)	.09 (48)	.08 (34)		.19 (5)	.29 (45)		.12 (4)	
Control Precision	.14 (25)	.14 (6)	.02 (4)	.04 (5)	.00 (2)		.16 (2)	.18 (8)		.11 (1)	
Rate Control	.12 (27)	.12 (3)	.12 (3)				.04 (3)	.10 (6)			
Finger Dexterity	.16 (32)	.20 (29)	.11 (42)	.15 (47)	.11 (25)	.12 (1)	.26 (8)	.09 (24)			.10 (2)
Manual Dexterity	.12 (31)	.24 (12)	.08 (34)	.17 (31)	.08 (14)	.21 (1)	.32 (8)	.10 (13)			.17 (2)
Wrist-Finger Speed	.05 (43)	.38 (17)	.18 (44)	.22 (34)	.24 (18)	.25 (1)	.26 (14)	.03 (16)	-.06 (1)		.05 (2)
Arm-Hand Steadiness	.16 (30)	.07 (5)	.00 (4)	-.02 (2)	.00 (3)		.25 (3)	.17 (7)			
Speed-of-Arm Movement	.11 (9)	.06 (1)	-.07 (1)				.06 (1)	.04 (2)			

NOTE: The top entry in each cell is the median correlation between the psychomotor and cognitive-perceptual abilities. The number in parentheses below is the number of correlations identified for this pair of abilities.

in any new selection and classification test battery.

Manual dexterity is the third psychomotor ability which appears to be moderately correlated with more than one cognitive-perceptual ability. Median correlations involving manual dexterity include .32 with perception ($N=8$ correlations), .24 with perceptual speed and accuracy ($N=12$ correlations), and .21 with memory ($N=1$ correlation). These median correlations are very similar to those between finger dexterity and these three cognitive-perceptual abilities. The correlations indicate that perceptual abilities are also relevant to manual dexterity tasks. The correlations are not so high, however, to eliminate manual dexterity from consideration as a predictor in the current research, especially given the evidence presented in Table 29 indicating the potential validity of manual dexterity measures for predicting training performance and job proficiency criteria in several job types.

The final psychomotor ability with moderate or high correlations with two or more cognitive-perceptual abilities is wrist-finger speed. Almost all of the correlations between wrist-finger speed and cognitive-perceptual abilities presented in Table 32 are based on studies with the GATB. Most of these correlations are quite high (median $r=.38$ with perceptual speed and accuracy, $N=17$ correlations; median $r=.26$ with perception, $N=14$ correlations; median $r=.25$ with memory, $N=1$ correlation; median $r=.24$ with number facility, $N=18$ correlations; and median $r=.29$ with reasoning, $N=34$ correlations).

One possible explanation for these high correlations is that wrist-finger speed is a measure of marking speed. In the GATB, one of the wrist-finger speed tests is simply a measure of the number of circles a subject can mark in a fixed period of time. To the extent that a subject's perceptual speed and accuracy, perception, memory, number facility, and reasoning test scores are affected by ability to mark an answer sheet quickly, it would be expected that correlations between these abilities and wrist-finger speed would be high.

The correlation of .38 with perceptual speed and accuracy is particularly easy to explain in this manner. For most perceptual speed and accuracy tests, subjects are presented with two character strings (e.g., names, addresses, nonsense syllables) and asked to determine whether the strings are exactly the same or different. In most cases, any differences between the two strings are only slight. Computer versions of this test demonstrate that most subjects can determine the correct answer to each item in just a fraction of a second. Indeed, it probably takes most subjects much more time to mark their answer sheets than to determine the correct answer. Thus, it is not surprising to see the high correlation between wrist-finger speed and perceptual speed and accuracy. The correlation almost certainly indicates that much of the variance in paper-and-pencil perceptual speed and accuracy tests is attributable to marking speed (i.e., wrist-finger speed) rather than to perceptual processes. The same could probably be said of most other highly speeded paper-and-pencil tests.

Therefore, the moderately high correlations between wrist-finger speed and cognitive-perceptual abilities can probably be attributed in large measure to test marking variance embedded in the cognitive-perceptual

ability tests. Since the Army is seriously considering changing to computerized administration of most of its selection and classification tests, it is unlikely that future tests of these cognitive-perceptual abilities will be so highly correlated with wrist-finger speed. Hence, it is difficult to justify excluding tests of wrist-finger speed from future selection and classification batteries simply on the basis of these correlations. Moreover, Table 29 shows that wrist-finger speed is a valid predictor of training performance and job proficiency in several job types, indicating that wrist-finger speed tests may have high utility for the Army.

One particularly interesting fact emerging from Table 32 is that correlations between psychomotor abilities and abilities which are highly related to general intelligence (e.g., verbal comprehension and reasoning) are generally quite low. For example, excluding wrist-finger speed, median psychomotor ability correlations with verbal comprehension range from $-.07$ for speed of arm movement ($N=1$ correlation) to $.12$ for rate control ($N=3$ correlations). These median correlations suggest that there is virtually no relationship between psychomotor abilities and g .

In sum, Table 32 provides strong evidence that the ability variance tapped by psychomotor ability tests is different from that tapped by cognitive-perceptual ability tests. Taken together, Tables 29 and 32 demonstrate that psychomotor ability tests can contribute unique variance to the prediction of job and training performance. This suggests that psychomotor ability testing may indeed have high utility for the Army.

Psychomotor Test Administration. Throughout most of the history of psychomotor testing, there have been two options for test administration format: paper-and-pencil measures and mechanical apparatus measures.

Psychologists, of course, have long been enamored with paper-and-pencil tests. Paper-and-pencil tests provide psychologists with the opportunity to control almost perfectly the presentation of test stimuli to subjects and to ensure that all subjects are provided with exactly the same test instructions and test conditions.

Much of this control is lost with apparatus tests. As a result, mechanical apparatus tests have several disadvantages in comparison with paper-and-pencil tests.

Perhaps the most serious disadvantage of mechanical apparatus tests is the lack of standardization. Two units of the same apparatus may be quite different with respect to actual operation. Springs may be new and resilient on one unit and old and unresponsive on the second. Nuts and bolts may be tight on one unit and loose on the second. Contact points on switches may be clean and new on one unit and old and corroded on the second. Such differences may seem trivial. Yet, when one considers that some of the more complicated mechanical apparatus consists of scores of springs, nuts, bolts, and switches, one begins to appreciate that the effects of these apparatus unit differences might be quite large indeed. Confirmation of this was provided by research conducted by the AAF during World War II (Melton, 1947). Researchers found numerous correlations between calibration measurements on apparatus and mean test scores on the various units of that apparatus.

A second major disadvantage of apparatus tests is that they frequently break down. This often results in missing or invalid data. In addition, since broken apparatus units are not as easy to replace as damaged or misprinted paper-and-pencil tests, broken apparatus units may play havoc with testing schedules.

Because of the problems associated with apparatus testing, psychologists have generally been very reluctant to use apparatus tests unless no alternative measures of the target abilities were available. As Melton, (1947) wrote regarding the development of apparatus tests by the AFF Aviation Psychology Program during World War II, "At no time was there the intention of measuring with apparatus some function which could be measured as meaningfully by the marking of an IBM answer sheet" (p. 54).

Nevertheless, as was noted previously in the taxonomy section, there appear to be several abilities (e.g., multilimb coordination, control precision, rate control, finger and manual dexterity) which defy measurement via paper-and-pencil tests (Cronbach, 1970). To assess these abilities, there appears to be no alternative to apparatus testing.

With respect to utility, the major issue is whether the problems associated with apparatus testing can be compensated by the gains in criterion-related validity resulting from the apparatus test scores. While they apparently never conducted any formal analysis of apparatus test utility, the military decided after World War II that apparatus testing was *not* worthwhile (Passey & McLaurin, 1966). This occurred in spite of data collected during World War II which demonstrated that the validity of many apparatus tests was quite high.

Research on the selection of helicopter pilots by Zeidner, Martinek, and Anderson (1958) and Fleishman's continuing research into the taxonomy of psychomotor abilities sparked new interest in apparatus testing in the late 1950s. One result of that interest was renewed research into the reliability of apparatus tests by Dobbins, Martinek, Anderson, and Rosenberg (1961). While Dobbins et al. were encouraged by the increments in predictive validity produced by psychomotor apparatus tests in the selection of Army helicopter pilots, they were generally pessimistic regarding the utility of such tests. In their words:

Results of the study reemphasized the need for caution in the use of apparatus tests. The study reaffirmed the fact that individual differences in machines do occur and that scores tend to fluctuate significantly with continued use of the same machine. A statistical solution may be the only feasible means of completely neutralizing apparatus differences. Ideally, however, score fluctuations through continuing machine use should be handled by proper maintenance and not by statistical conversion. (Dobbins et al., 1961, p. 9)

The need for continuous apparatus maintenance, as well as the need to conduct all testing through centralized facilities, added greatly to the expense associated with apparatus testing and largely nullified any gains in utility associated with these tests.

In spite of these consistent negative findings with respect to appar-

atus test utility, psychologists involved in pilot selection continued to be interested in including apparatus tests in their selection batteries. For example, reviews of the psychomotor ability literature by Passey and McLaurin (1966), Hunter (1975), and Imhoff and Levine (1981) all resulted in the recommendation that apparatus tests be included in future Air Force selection and classification batteries.

Events with perhaps the greatest impact on apparatus testing, however, occurred during the 1970s. Breakthroughs in microprocessor technology resulted in a new generation of solid state apparatus tests.

Among the first researchers to experiment with computerized apparatus tests were Sanders, Valentine, and McGrevey (1971). They developed new versions of the Two-Hand Coordination Test and the Complex Coordination Test. (See Appendix A for pictures and descriptions of both the original mechanical and the new computerized versions of these two tests.) For both tests, Sanders et al. used computer graphics video displays to present the test stimuli to subjects. Subjects were then required to perform a tracking task using either a pair of joy sticks (for the Two-Hand Coordination Test) or a joy stick and a rudder control (for the Complex Coordination Test). Sanders et al. estimated the parallel forms reliability to be .81 for the Two-Hand Coordination Test and .92 for the Complex Coordination Test. Sanders et al. also found that the tests did not correlate highly with any of the cognitive-perceptual tests or composites in the Air Force Officer Qualifying Test (AFOQT). In a follow-up to this initial study, McGrevey and Valentine (1974) found that the Complex Coordination Test was extremely effective as a predictor of graduation from undergraduate pilot training ($r=.44$, $N=121$, $p<.05$). By contrast, a regression-derived prediction equation involving all 13 of the tests comprising the AFOQT yielded a multiple correlation of only .37. The predictive validity of the Two-Hand Coordination Test was not quite as high as that of the Complex Coordination Test ($r=.18$, $N=121$, ns).

Since the pioneering research of Sanders, Valentine, and McGrevey, a number of other psychomotor tests have been adapted for the computer. In some cases, these tests are rather direct measures of Fleishman's multilimb coordination, rate control, and control precision abilities (e.g., the Complex Coordination Test developed by Myers et al., 1982). In other instances, the tests are rather elaborate job simulations designed to predict performance in specific jobs (e.g., the tests using the Willey "Burst-on-Target" Simulator developed by Biers & Sauer, 1982, which are used to predict tank crew performance). Table 33 contains a list of some of the computerized psychomotor tests currently being used and studied by the U.S. military. It should be noted that none of the tests listed in Table 33 are measures of manual or finger dexterity. To date, no one has succeeded in producing a computerized dexterity test. As a result, there is currently no alternative to traditional apparatus measures for these two abilities.

Computerized apparatus tests have several advantages over mechanical apparatus tests.

First, the computerized tests tend to be much more reliable. This is because the solid state components used in these apparatus are much less affected by wear and tear than the mechanical parts used in the older

Table 33

Computerized Psychomotor Tests in Use in the Military

Test	Location	Machine
Complex Coordination (stick and rudder)	AFHRL	Analog apparatus
Complex Coordination (two tracking tasks)	NAMRL	Analog apparatus ^a
Computerized Target Engagement	Ft. Knox	Apple II
Critical Tracking	NAMRL	PDP-11
Gunner Tracking Task (using the Willey Burst-on-Target Simulator)	Ft. Knox	Analog apparatus
Helicopter Simulator	Ft. Rucker	Analog apparatus
One-Dimensional Compensatory Tracking	NAMRL	PDP-11
Perceptronics Simulator	Ft. Knox	Analog apparatus
Psychomotor Tracking Task	Ft. Knox	Apple II
Tank Turret Simulator	Ft. Knox	Analog apparatus
Tank Video Game	(b)	Apple II
Target Acquisition Task (using the Willey Burst-on-Target Simulator)	Ft. Knox	Analog apparatus
Two-Dimensional Compensatory Tracking	Ft. Rucker	Apple II
Two-Hand Coordination	AFHRL	Analog apparatus

a. NAMRL is currently adapting this to an Apple II computer with joy stick and foot pedals.

b. This test is being developed under a contract with ARI. The work is being carried out at NAMRL.

apparatus tests. Thus, there is much less intra-unit and inter-unit variability among the computerized tests than there was among the mechanical tests. This is not to say that the computerized tests are perfectly standardized. No two circuit boards will ever be exactly alike, so there will always be some test score variance attributable to individual differences in apparatus units. In addition, even the most state-of-the-art computerized apparatus test still requires many mechanical parts. For example, almost all tracking and coordination tests require the subject to use mechanical controls such as joy sticks or control sticks to respond. To date, there is very little in the published psychological literature which indicates how much test score variance is attributable to differences in these mechanical controls, but the magnitude of this variance is likely to be significant. As the use of computerized psychomotor tests increases over the next several years, it is hoped that research on inter-unit differences among computerized apparatus tests will be conducted and published.

A second advantage of computerized apparatus tests over mechanical apparatus tests is that computerized tests are much less likely to break down. This is because computers have fewer moving parts than the old mechanical apparatus. As a result, maintenance costs and costs associated with "down time" have been reduced considerably with computerized testing.

Third, computerized testing results in even greater standardization in testing conditions than paper-and-pencil testing. With computerized testing, instructions to subjects can be presented on the video display terminal, a standard set of practice trials can be conducted, and the test can then be initiated by the subject. No administrator intervention is necessary. Indeed, most new batteries of computerized tests automatically proceed from one test to the next without any intervention by the test administrator. Thus, computerized testing helps eliminate almost all of the variance in test scores attributable to nonstandardized testing conditions.

Finally, computerized testing greatly reduces data capture and scoring costs. Computers can be programmed to automatically record each response made by the subject and to score these responses at the conclusion of the test. This procedure effectively eliminates all data capture errors (e.g., data which are keypunched incorrectly).

In sum, the advent of computerized psychomotor testing has eliminated many of the concerns associated with the unreliability and expense of mechanical apparatus testing. In fact, computerized psychomotor tests may actually be cheaper to administer and score than traditional paper-and-pencil cognitive ability tests. For this reason, and also because psychomotor apparatus tests often contribute unique variance to the prediction of performance in many jobs, the utility of psychomotor testing appears to be substantial.

Psychomotor Testing and Group Differences

Over the past two decades, a great deal of research has been directed toward examining group differences in ability test scores. Much of this research has been motivated by concerns over test fairness and the legal requirements of selection testing. (A summary of the legal and related

psychometric issues in selection testing was presented in the literature review for cognitive abilities. Readers interested in differential validity, models of test fairness, and major legal decisions related to selection testing are referred to this summary.) Unfortunately, to date, almost all of this research has been conducted in the cognitive-perceptual and physical ability domains. As a result, few data are available on group differences in psychomotor abilities. In the literature reviewed for this report, only five studies discussed group differences and/or differential validity. Results from these studies are discussed below.

The GATB and Group Differences. Data on group differences in GATB psychomotor aptitude scores were reported in a study by Fozard and Nuttall (1971) and in the GATB test manual (U.S. Department of Labor, 1970).

Fozard and Nuttall were interested in differences in GATB aptitude scores across age groups and socioeconomic status (SES) groups. Subjects in their study were 1,146 men participating in a longitudinal study of aging. At the time of testing, the men ranged from 28 to 83 years old. SES was estimated for each subject using the Warner system (Warner, Meeker, & Eels, 1960). With the Warner system, SES estimates are based primarily on the subject's occupation (or occupation at the time of retirement). Even though the subjects were participating in a longitudinal study, for this particular study each subject was tested only once (i.e., a cross-sectional design was used). This means that the subjects in the 25-35 year old age group were not the same subjects as those in the 46-50 year old age group. Consequently, some of the age differences in psychomotor test scores identified during this research may be attributable to generational differences rather than to biological changes associated with aging.

For their analyses, Fozard and Nuttall performed two-way (age x SES) analyses of variance for each of the nine GATB aptitude areas. They found statistically significant age effects for all nine aptitudes ($p < .001$ for all three psychomotor aptitudes). In addition, they found significant SES effects for Finger Dexterity ($p < .01$), Motor Coordination ($p < .001$), and all six cognitive-perceptual aptitudes ($p < .001$ for all six). Manual Dexterity was the only aptitude which failed to show a significant effect for SES. Finally, Fozard and Nuttall tested for age by SES interactions. No statistically significant interactions were found for any of the nine aptitudes. Table 34 provides a summary of the percentage of variance attributable to age and SES for each of the three psychomotor aptitudes.

Fozard and Nuttall also examined the probable impact of age and SES on selection practices. Typically, GATB selection decisions are based on Occupational Aptitude Patterns (OAPs). The OAP for an occupational group is based on the two, three, or four aptitudes correlated most highly with job or training performance criteria. Cutoff scores are established for each of these aptitudes. Applicants are deemed qualified for an occupational group if their aptitude scores equal or exceed the cutoffs for all of the aptitudes comprising the OAP. Because of the negative correlation between age and GATB aptitude scores, Fozard and Nuttall found that age was strongly related to the number of occupational groups for which applicants qualified. For example, Fozard and Nuttall compared the average aptitude profiles of the oldest and youngest groups of subjects in the highest SES group. The average profile of the youngest subjects qualified them on all 36 OAPs, but the oldest subjects qualified on only 17 OAPs. Within the

Table 34

Percentage of Variance Attributable to Age
and Socioeconomic Status for the
Three GATB Psychomotor Scales

Aptitude Area	<u>Percentage of Variance Attributable to:</u>	
	Age	Socioeconomic Status
Motor Coordination	3%	5%
Finger Dexterity	14%	1%
Manual Dexterity	11%	0%

Note. The data are from "General Aptitude Test Battery scores for men differing in age and socioeconomic status" by J. L. Fozard and R. L. Nuttall, 1971, *Journal of Applied Psychology*, 55, p. 375.

lowest SES group, the youngest group of subjects qualified on 28 OAPs. In contrast, the oldest subjects in this SES group qualified on only nine OAPs. Fozard and Nuttall concluded, "Clearly, the older worker in all SES groups is severely handicapped when compared to younger men on the basis of OAP cutoff scores" (Fozard & Nuttall, 1971, pp. 376-377).

Sex differences on the GATB psychomotor scales were studied in a sample of over 7,000 high school students (U.S. Department of Labor, 1970). Significant differences were found for two of the psychomotor scales. Boys scored .4 standard deviation higher than girls on the Motor Coordination scale and .35 standard deviation higher than girls on the Finger Dexterity scale. No significant sex differences were found for the Manual Dexterity scale.

Since selection decisions with the GATB are typically based on OAPs, no data are available to compare the criterion-related validity of the GATB psychomotor scales for males and females. The GATB manual does, however, report separate criterion-related validities for males and females for 11 different OAPs, six of which include one or more psychomotor scales. For these six OAPs, there is little evidence of differential validity for males and females. Based on these scant data, it appears as though the GATB psychomotor scales are equally valid for males and females.

Table 35 summarizes the data on ethnic group differences in GATB psychomotor aptitude scores (U.S. Department of Labor, 1970).

The sample of job applicants tested in local California employment offices is probably most representative of the general population. For all three psychomotor scales, overall differences between minority and nonminority applicants is less than .1 standard deviation. Blacks scored lowest of all groups on all three psychomotor scales, while Hispanics and Orientals scored higher than nonminority applicants on all three scales. Most group differences were quite small, however.

The pattern of group differences was similar in the samples of welders and draftsmen. Nonminority welders scored between .4 and .6 standard deviation higher than minority welders on the three psychomotor scales. All but one of these minority welders was black. In the first sample of draftsmen, minority draftsmen--over 90% of whom were Hispanic or Oriental--scored approximately .5 standard deviation higher than nonminority draftsmen on all three psychomotor scales. Hispanics and Orientals also scored higher than nonminorities in the second sample of draftsmen. While separate means are not available for American Indians and blacks in this sample, the data certainly suggest that blacks scored lower than Hispanics, Orientals, and nonminorities on all three scales.

Unfortunately, there are no data available to evaluate whether the psychomotor scales are differentially valid for minorities and nonminorities.

In summary, the group differences data available for the GATB show significant age and ethnic group differences for all three psychomotor scales and significant sex differences for Motor Coordination and Finger Dexterity. Since the data on age differences were cross-sectional rather than longitudinal, however, and since no data were available on the

Table 35

Ethnic Group Differences in GATB Psychomotor Aptitude Scores^a

Psychomotor Aptitude Area	Ethnic Group	Sample		
		Job Applicants Tested in Local Employment Offices in California ^c	Production Line ^d Welders	Draftsmen ^e Draftsmen ^f
Motor Coordination ^b	Minorities	101.2	81.0	115.8
	American Indians	104.7		112.6 (97.8) ^g
	Blacks	98.4		
	Hispanics	102.7		110.8
	Orientals	110.0		121.3
	Nonminorities	100.2	93.5	104.7 104.5
Finger Dexterity	Minorities	93.7	77.4	99.1
	American Indians	94.5		(95.4) ^g
	Blacks	89.2		
	Hispanics	97.7		98.8
	Orientals	97.6		101.1
	Nonminorities	95.6	88.2	89.3 94.9
Manual Dexterity	Minorities	104.1	84.7	108.3
	American Indians	112.6		(97.7) ^g
	Blacks	99.8		
	Hispanics	107.1		109.1
	Orientals	107.6		112.3
	Nonminorities	102.8	92.3	97.9 101.4

Note. The data are from Manual for the USIES General Aptitude Test Battery by the U. S. Department of Labor, Manpower Administration, 1970, Washington, DC: U. S. Government Printing Office.

a. GATB scales have a mean of 100 and a standard deviation of 20 in the general population.

b. Motor Coordination is a measure of wrist-finger speed.

c. The sample for this study included 3,145 minority applicants (171 American Indians, 1,413 blacks, 1,425 Hispanics, and 136 Orientals) and 6,672 nonminority applicants.

d. The sample for this study included 57 minority welders (1 American Indian and 56 blacks) and 59 nonminority welders.

e. The sample for this study included 55 minority draftsmen (5 blacks, 29 Hispanics, and 21 Orientals) and 177 nonminority draftsmen.

f. The sample for this study included 89 minority draftsmen (7 American Indians, 9 blacks, 38 Hispanics, and 35 Orientals) and 333 nonminority draftsmen.

g. Separate mean aptitude scores were not reported for American Indians or blacks. Therefore, the numbers in parentheses represents the combined black-American Indian mean aptitude scores.

criterion-related validity of the GATB for different age groups, it is difficult to determine whether the GATB is biased in favor of younger subjects. Similarly, the presentation of validity data for males and females and for various ethnic groups makes it impossible to determine whether the GATB is a valid predictor for each of these groups and whether the observed group differences in test scores are actually related to job and/or training performance.

Group Differences and the ACB. During World War II, the United States armed forces were, for the most part, segregated by race and sex. Blacks trained and fought in separate units from whites, and women were not allowed to hold combat positions. Nevertheless, large numbers of blacks and women were needed and used as pilots during the war. Black units in the AAF played a vital role in combat in Europe and the Pacific, and women assigned to the Women's Auxiliary Service Pilots (WASPs) were responsible for shuttling troops, supplies, and equipment overseas.

The Aviation Psychology Program research team administered the ACB to samples of black student pilots and WASPs during World War II in an effort to determine the validity of this battery for special samples of aviation students. Results for four ACB psychomotor tests (Melton, 1947) are summarized in Tables 36 and 37.

Table 36 reveals that white student pilots scored higher than black student pilots on all four psychomotor tests. Differences in mean scores ranged from .17 standard deviation on the Santa Ana Finger Dexterity Test to .44 standard deviation on the Two-Hand Coordination Test. On the three tests of multilimb coordination, the mean black-white differences was .29 standard deviation. These black-white differences are quite a bit smaller than those reported for cognitive-perceptual abilities in that literature review and may be of little or no practical significance.

Of special interest are the validities of these psychomotor tests in the black sample. While median validities for the three multilimb coordination tests ranged from .24 to .26 across several samples of white student pilots, validities in the black sample ranged from only .04 to .14. (The Santa Ana Finger Dexterity Test was not a valid predictor of graduation from elementary pilot training in either group.) The researchers offered "no explanation of the failure of the test(s) in the case of the negro students....(N)one of the apparatus tests showed appreciable validity" (Melton, 1947, p. 142). This is the only example of differential validity of psychomotor tests for blacks and whites uncovered in the literature reviewed.

An important question is whether this finding of differential validity generalizes to other psychomotor abilities, other jobs, and other samples of blacks and whites. The student pilots who completed the ACB represented a rather select group. Service members who wished to enter flight training first had to pass the AAF Qualifying Examination. This examination was comprised of several cognitive ability tests, including verbal aptitude, knowledge of current affairs, mechanical aptitude, mathematics, practical judgment, and interpretation of data (Davis, 1947). Miller (1947) reported that those passing the AAF Qualifying Examination represented the top 25% of all high school graduates. Unfortunately, Davis did not report any data on black-white differences for the examination. If, however, one assumes a

Table 36

**Descriptive Statistics and Validities^a for Selected Psychomotor Tests
in the Aircrew Classification Battery for Samples of Black and White Aviation Students**

Psychomotor Test ^b	Ethnic Group																
	Blacks										Whites						
	N	P _g	C	M _g	d	M _e	M _e	f	M _o	r _{bis}	g	N	P _g	M _g	M _e	M _o	r _{bis}
Complex Coordination	658	.68		50.76		50.18		50.57		.04		40,527	.81	53.64	48.40	52.64	.26
Two-Hand Coordination	657	.67		48.80		47.26		48.29		.10		24,124	.80	53.40	50.10	52.74	.24
Rudder Control	657	.67		53.78		51.74		53.11		.14		10,278	.78	55.96	53.48	55.41	.24
Santa Ana Finger Dexterity	660	.67		49.24		50.18		49.55		-.06		19,668	.81	51.40	50.62	51.25	.07

NOTE. The data are from Apparatus tests (Army Air Forces Aviation Psychology Research Report No. 4) by A. V. Melton (Ed.), 1947, Washington, DC: U. S. Government Printing Office.

a. The criterion for all validity assessments was graduation or elimination from elementary pilot training.

b. All tests had a mean of 50 and a standard deviation of 10 in the target population.

c. P_g represents the proportion of candidates who graduated from elementary pilot training.

d. M_g represents the mean test score of the students who graduated from training.

e. M_e represents the mean test score of the students who were eliminated from training.

f. M_o represents the overall mean test score, including both training graduates and eliminatees.

g. r_{bis} represents the biserial correlation between students' test scores and graduation from training. For white students,

r_{bis} is the median correlation across several classes of students.

Table 37

**Descriptive Statistics and Validities^a for Selected Psychomotor Tests
in the Aircrew Classification Battery for Samples of Male and Female Aviation Students**

Psychomotor Test ^b	Sex											
	Females						Males					
	N	p _g	M _d _g	M _e	M _f	r _{bis} ^g	N	p _g	M _d _g	M _e	M _o	r _{bis}
Complex Coordination	191	.71	47.72	43.00	46.35	.22	10,143	.71	52.96	48.40	51.64	.33
Two-Hand Coordination	191	.71	39.66	33.98	38.01	.34	7,160	.78	53.10	48.06	51.99	.29
Rudder Control	191	.71	62.82	59.54	61.87	.32	920	.61	55.00	45.50	51.30	.55
Santa Ana Finger Dexterity	191	.71	47.42	48.86	47.84	-.04	7,165	.78	51.58	50.04	51.24	.08

Note. The data are from Apparatus tests (Army Air Forces Aviation Psychology Research Report No. 4) by A. W. Melton (Ed.), 1947, Washington, DC: U. S. Government Printing Office.

- a. The criterion for validity assessments for female students was graduation from advanced pilot training. Although it is not completely clear from Melton's report, it appears that these women were tested after they completed elementary and basic training but before they began advanced training. The criterion for validity assessments for male students was also graduation or elimination from advanced pilot training. The males, however, were tested prior to elementary pilot training. Thus, male eliminees included students eliminated from elementary and basic training as well as students eliminated from advanced training.
- b. All tests had a mean of 50 and a standard deviation of 10 in the general population tested.
- c. p_g represents the proportion of candidates who graduated from advanced pilot training.
- d. M_g represents the mean test score of the students who graduated from training.
- e. M_e represents the mean test score of the students who were eliminated from training.
- f. M_f represents the overall mean test score, including both graduates and eliminees.
- g. r_{bis} represents the median biserial correlation between students' test scores and graduation from training across several training classes.

black-white difference of one standard deviation for AAF Qualifying Examination scores (see the summary of black-white cognitive ability differences in that literature review), then only about 10% of the black population would have been able to pass the examination.

One result of this screening with the AAF Qualifying Examination is that validity coefficients for any subsequent screening tests could possibly be severely attenuated for blacks. This may explain some of the differences in validity coefficients for psychomotor tests for blacks and whites. Since correlations between multilimb coordination and cognitive ability tests are quite low, however (see Table 32), there was probably not a great deal of range restriction in multilimb coordination test scores for the student pilot candidates who completed the ACB (vis-a-vis the general population). Therefore, range restriction cannot explain all of the differences in these validity coefficients. Moreover, the proportion of entering student pilots who actually graduated from pilot training was approximately .80 for whites and .67 for blacks. Since correlations between dichotomous and continuous variables are attenuated as the split on the dichotomous variable departs from .50-.50, the validity coefficients for whites are actually attenuated slightly more than those for blacks.

A comparison of mean ACB psychomotor test scores for males and females is even more difficult to interpret because of differences in the sampling of these two groups.

While it is not completely clear from the description provided by Melton (1947), it appears that the women tested were WASPs who had graduated from elementary and basic training and who were preparing to enter advanced training. As noted previously, the AAF Qualifying Examination eliminated all but the top 25% of high school graduates from pilot training. Typically then, 35-40% of those who did pass the AAF Qualifying Examination were eliminated prior to advanced training (Miller, 1947). These eliminations were generally the result of flight deficiencies which occurred during elementary and basic training. Thus, the WASPs involved in this research had already soloed successfully during elementary training and demonstrated proficiency on a number of basic flight maneuvers during basic training. If one assumes, based on validity results presented for male student pilots in Table 9, that multilimb coordination is a valid predictor of graduation from elementary flight training for women (cf. Table 36), then the WASPs tested for this research probably scored higher on tests of multilimb coordination than females in general. Therefore, the correlations between multilimb coordination and graduation from advanced training for women reported in Table 37 should be somewhat attenuated due to range restriction. In addition, the mean test scores reported in the table are probably an overestimate of the mean scores for women in general.

The males, whose data are reported in Table 37, were tested prior to elementary flight training. Like their female counterparts, these men were screened with the AAF Qualifying Examination prior to testing. Unlike the WASPs, though, these men had not yet demonstrated their flight aptitude in elementary or basic training. Since the cognitive screening tests used in the AAF Qualifying Examination are not highly correlated with multilimb coordination, there is probably not a great deal of range restriction with respect to multilimb coordination for these men.

Given these sampling differences, one might expect the following:

1. Mean scores on tests of multilimb coordination will be slightly higher for females than males. This is because WASPs scoring low on these tests would be more likely to attrite during elementary and basic flight training than WASPs scoring high on these tests. Male student pilots, on the other hand, would not have had an opportunity to attrite during elementary or basic training prior to testing.
2. Multilimb coordination validity coefficients for females will be attenuated relative to those for males. This is because the range of scores for females on multilimb coordination tests will be reduced due to the elevated attrition rate during elementary and basic training for WASPs with relatively low standing on this ability.

Table 37 shows that neither of these predictions was upheld. First, males scored .53 standard deviation higher than females on the Complex Coordination Test and 1.40 standard deviations higher than females on the Two-Hand Coordination Test. Females, however, scored 1.06 standard deviations higher than males on the Rudder Control Test. Yet all three of these tests are measures of multilimb coordination. One possible explanation for the large female advantage on the Rudder Control Test is that the behaviors required on this apparatus test are very similar to those required in flying a plane. The females tested all had previous flight experience, while none of the males did. Melton (1947) reported a correlation of .68 between previous flying experience and scores on the Rudder Control Test. Thus, this experience may account for the high mean score for females on this test. Nevertheless, from the data presented in Table 37, it is extremely difficult to determine whether there are sex differences in mean level of multilimb coordination ability. Second, the validity coefficients for multilimb coordination are very similar for females and males. On the average, the validities appear to be just slightly higher for males.

Data in Table 37 also indicate that males scored .34 standard deviation higher than females on the Santa Ana Finger Dexterity Test. The test was not a valid predictor of training attrition for either group.

In sum, the data on group differences from the AAF Aviation Psychology Program pose more questions than they answer. White males scored slightly higher than black males on three different measures of multilimb coordination. None of the tests was a valid predictor of attrition from elementary pilot training for blacks, while all three tests were valid predictors in white student pilot samples. Sex differences in scores on multilimb coordination tests were more pronounced than ethnic group differences. Unfortunately, the direction of these sex differences was inconsistent. Males scored considerably higher than females on two multilimb coordination tests, but females scored quite a bit higher than males on a third test of this ability. Additional analyses of these data indicated no significant differences in test validities for males and females. Finally, on the Santa Ana Finger Dexterity Test, white males scored slightly higher than black males, who in turn scored slightly higher than females. The test was not a valid predictor of graduation from flight training for any of the groups.

Group Differences and the Cross-Cultural Aircrew Aptitude Battery CCAAB). The CCAAB was developed to aid in the selection of foreign students receiving aircrew training from the U.S. Air Force (Mullins, Keeth, & Riederich, 1968). At the time, many foreign students were coming to this country for aircrew training under the auspices of the Military Assistance Program (MAP). The goal of MAP was to provide technical and pilot training to members of the armed forces from allied countries.

One problem with MAP was that different countries used different selection procedures when choosing students to participate in the program. To help solve this problem, the Air Force decided "to develop a cross-cultural, cross-language battery of tests which could be used by the individual countries, regardless of language and cultural differences, to improve their selection procedures and aid in predicting success of those chosen for pilot or technical training" (Mullins et al., p. 1). The initial CCAAB was comprised of 27 tests, including seven paper-and-pencil and three apparatus psychomotor tests.

Results from the initial validity study for pilots were quite promising. Subjects were 120 pilot trainees from a range of foreign countries. Two of the apparatus tests, Rudder Control and Complex Coordination, correlated significantly with graduation from pilot training.

In a subsequent follow-up study, the CCAAB was administered to a sample of 244 South Vietnamese pilot trainees (Croll et al., 1973). Because the CCAAB tests were administered by Vietnamese personnel in Vietnam, Croll et al. decided that it would be too complicated for the administrators to oversee the apparatus tests. Therefore, only the paper-and-pencil tests were administered. Two of the seven paper-and-pencil psychomotor tests correlated more than .20 with graduation from pilot training. Line Control (a measure of arm-hand steadiness) correlated .24 with graduation, while Trace Tapping II (a measure of aiming) correlated .29 with graduation.

In the first of these two CCAAB studies, Mullins et al. (1968) did not score their apparatus tests in the same measurement units which previous researchers had used. Therefore, it was not possible to compare mean apparatus test scores for these foreign student pilots with mean apparatus test scores for other groups. Nor was it possible to compare the performance of these subjects on the paper-and-pencil psychomotor tests with the performance of other groups of subjects. This was because the paper-and-pencil psychomotor tests were all new measures which the researchers developed specifically for the CCAAB.

It was possible, however, to compare the validities Mullins et al. (1968) obtained for the apparatus tests with validity results which Melton (1947) reported from the AAF Aviation Psychology Program during World War II. These results are summarized in Table 38. Validities for Complex Coordination and Rudder Control, both of which are measures of multilimb coordination, were approximately equal in the two groups. The validities for the control precision test Rotary Pursuit, however, were quite different (median $r = .23$ for American student pilots, $r = -.01$ for foreign student pilots). While these data suggest that Rotary Pursuit may be a differentially valid predictor of graduation from pilot training for these two groups, such conclusions must be tentative until additional validity data

Table 38

Validity of Three Psychomotor Tests for Predicting Graduation
from Pilot Training for Foreign and American Student Pilots

Psychomotor Test	Subjects	
	Foreign Student Pilots Training with the U. S. Air Force in 1968 ^a	American Student Pilots Training with the U. S. Army Air Forces during World War II ^b
Complex Coordination	.21	.26
Rudder Control	.38	.24
Rotary Pursuit	-.01	.23

- a. The data in this column are from *Selection of foreign students for training in the United States Air Force* (AFHRL-TR-68-111) (p. 7) by C. J. Mullins, J. B. Keeth, and L. D. Riederich, 1968, Lackland Air Force Base, TX: U. S. Air Force Human Resources Laboratory, Personnel Research Division. The validity coefficients reported in this table are mean validities over five trials for Complex Coordination, six trials for Rudder Control, and three trials for Rotary Pursuit. $N=120$ for all tests and trials. The criterion was graduation from pilot training.
- b. The data in this column are from *Apparatus tests* (Army Air Forces Aviation Psychology Program Research Report No. 4) by A. W. Melton (Ed.), 1947, Washington, DC: U. S. Government Printing Office. The validity coefficients reported in this table are median validities across several classes of student pilots. $N=40,527$ for Complex Coordination (18 classes of student pilots), $N=10,278$ for Rudder Control (six classes of student pilots), and $N=8,955$ for Rotary Pursuit (seven classes of student pilots). The criterion was graduation from elementary pilot training.

for foreign student pilots are collected.

Summary of Group Differences Data. Table 39 summarizes the data on group differences for each of the nine major psychomotor abilities. No data are available for four of these abilities: rate control, aiming, arm-hand steadiness, and speed of arm movement. In addition, there are no data on average group tests score differences for control precision.

Mean scores for whites on tests of multilimb coordination, finger dexterity, manual dexterity, and wrist-finger speed are typically higher than those for blacks. Differences in mean scores range from .2 to .5 standard deviation. These differences are quite a bit smaller than the average black-white difference of one standard deviation for cognitive ability tests reported in the previous chapter. Additional data from the GATB indicate that Hispanics and Orientals score higher than whites on all three psychomotor scales. Average differences range from .4 standard deviation on the Finger Dexterity scale to 1.3 standard deviations on the Motor Coordination (i.e., wrist-finger speed) scale. American Indians also score higher than whites on two scales, the Manual Dexterity scale (.7 standard deviation) and the Motor Coordination scale (.3 standard deviation).

Data on sex differences show that males score higher than females on tests of finger dexterity (.35 standard deviation) and wrist-finger speed (.4 standard deviation). There are no sex differences in mean scores on tests of manual dexterity. Large sex differences have been found on three tests of multilimb coordination. The direction of these differences is inconsistent, however. In research conducted during World War II, males scored an average of one standard deviation higher than females on two tests of multilimb coordination while females scored slightly more than one standard deviation higher than males on a third multilimb coordination test.

Very little information is available regarding the differential validity of psychomotor tests. The available validity data suggest that tests of multilimb coordination, finger dexterity, manual dexterity, and wrist-finger speed are equally valid predictors of job and training performance for males and females. Two possible instances of differential validity were uncovered, however. First, tests of multilimb coordination had no validity as predictors of graduation from training for black student pilots. Ironically, these tests are the most valid psychomotor predictors of graduation from flight training for white male and female student pilots. Second, Rotary Pursuit did not predict graduation from training in a sample of foreign student pilots. For white American student pilots, however, the average validity of this control precision test was .23 across seven different samples.

The Use of Single- vs. Multiple-Construct Predictors

One of the major decisions facing any selection researcher is whether to use "factor-pure" single-construct ability measures or factorially complex multiple-construct work sample measures as performance predictors.

Researchers who are primarily interested in understanding the abilities involved in performing a job often prefer to use well-established tests and

Table 39

Summary of Group Differences Data for Nine Psychomotor Abilities

Psychomotor Ability	Summary of Group Differences in Average Test Scores	Summary of Differential Validity Evidence
Multilimb Coordination	White male student pilots scored slightly higher than black male student pilots on three different tests, but the average difference was only .29 standard deviation. Male-female differences were inconsistent. White male student pilots scored an average of .96 standard deviations higher on two tests of this ability, but females scored 1.06 standard deviations higher on a third test.	Tests were equally valid predictors of graduation from pilot training for white males, white females, and male foreign student pilots, with validities ranging from .22 to .55 across various samples. Tests were uncorrelated with graduation from pilot training for black males, however, with validities for three tests ranging from .04 to .14.
Control Precision	No data available.	Rotary Pursuit was a valid predictor of graduation from pilot training for white American males (median $r = .23$ across seven classes of pilot trainees), but had no validity for foreign student pilots (mean $r = -.01$ across three trials).
Rate Control	No data available.	No data available.

(Continued)

Table 39 (Continued)

Summary of Group Differences Data for Nine Psychomotor Abilities

Psychomotor Ability	Summary of Group Differences in Average Test Scores	Summary of Differential Validity Evidence
Finger Dexterity	<p>White male student pilots scored approximately .2 standard deviation higher than black male student pilots on the Santa Ana Finger Dexterity Test. In turn, black male student pilots scored approximately .2 standard deviation higher than female student pilots. Data for several samples from the GATB Finger Dexterity scale showed that Hispanics and Orientals scored an average of .4 standard deviation higher than whites, while whites scored an average of .5 standard deviations higher than blacks. There was no difference in the average scores of whites and American Indians. In a sample of 7,000 high school students, boys averaged .35 standard deviation higher than girls on this GATB scale. Other GATB research has demonstrated that older subjects and subjects from lower SES groups score lower on this scale than younger subjects and subjects from higher SES groups, respectively.</p>	<p>The Santa Ana Finger Dexterity Test was not a valid predictor of graduation from pilot training for blacks, whites, or women. Data from the GATB provide no evidence of differential validity of the GATB Finger Dexterity scale for males and females.</p>

(Continued)

Table 39 (Continued)

Summary of Group Differences Data for Nine Psychomotor Abilities

Psychomotor Ability	Summary of Group Differences in Average Test Scores	Summary of Differential Validity Evidence
Manual Dexterity	Hispanics, Orientals, and American Indians scored an average of .7 standard deviation higher than whites on the GATB Manual Dexterity scale. Whites in turn scored an average of .4 standard deviation higher than blacks on this scale. Data from a large sample of high school students showed no sex differences on this GATB scale. Additional analyses with GATB data revealed significant SES effects but no significant age effects.	There was no evidence of differential validity of the GATB Manual Dexterity scale for males and females. No data were available to evaluate the differential validity of this scale for minorities and nonminorities.

(Continued)

Table 39 (Continued)

Summary of Group Differences Data for Nine Psychomotor Abilities

Psychomotor Ability	Summary of Group Differences in Average Test Scores	Summary of Differential Validity Evidence
Wrist-Finger Speed	Orientals averaged almost 1.3 standard deviations higher than whites on the GATB Motor Coordination scale, while Hispanics and American Indians averaged .3 standard deviations higher than whites. Blacks scored an average of .5 standard deviations lower than whites on this scale. In a large sample of high school students, boys averaged .4 standard deviations higher than girls on this scale. Additional GATB data show that scores on this scale are negatively correlated with age and positively correlated with SES.	There was no evidence of differential validity of the GATB Motor Coordination scale for males and females. No data were available to evaluate the differential validity of this scale for minorities and nonminorities.
Aiming	No data available.	No data available.
Arm-Hand Steadiness	No data available.	No data available.
Speed of Arm Movement	No data available.	No data available.

constructs for ability measures. In this manner, they can specify the abilities underlying job performance with more confidence based on their knowledge and understanding of the test and ability construct they have elected to measure. In addition, they can compare ability requirements for different jobs to determine the abilities common to performance within and between jobs or job families.

Conventional wisdom, however, suggests that the best way to predict job performance is to use predictors which are similar in content to the job (i.e., predictors which are "samples" of performance; Wernimont & Campbell, 1968). Researchers who favor this approach note that performance on most jobs requires a complex blend of knowledges, skills, and abilities, and that single-construct predictors fail to tap this complexity.

In reality, there is no answer to the question of whether factor-pure single-construct measures or factorially complex multiple-construct measures are better predictors. This is because there is no such thing as a factor-pure single-construct measure. Previously, in the discussion of the taxonomy of the psychomotor domain, it was shown that abilities vary along a continuum of complexity. Within the psychomotor domain, for example, the psychomotor abilities identified by Fleishman actually consist of a number of more elementary perceptual and motor abilities. It was also suggested that the abilities comprising Fleishman's psychomotor taxonomy could in turn be grouped into more complex abilities, and that ultimately one could probably identify a general psychomotor coordination or "klutziness" ability which would be analogous to general intelligence (i.e., *g*) within the cognitive domain. Finally, it was concluded that the appropriate level of complexity for an ability taxonomy for any selection research is dependent upon the complexity of the criteria one is attempting to predict. That is, if one were attempting to predict performance on tasks which require a number of perceptual and psychomotor skills and abilities, the best course would be to choose predictors which were similar to those tasks and which would thus tap those required skills and abilities. Conversely, if the ability requirements of a task were quite simple, then measures of more basic processes would likely be the best performance predictors.

This approach to the selection of predictors is not unlike that described in the section on content validity and content-oriented test development. There again it was suggested that tests which are similar in content to the actual tasks performed on the job (i.e., samples of performance) would be the most valid predictors of performance.

Jobs which are heavily dependent upon psychomotor (vs. cognitive) abilities would seem to be particularly good candidates for multiple-construct work sample selection tests. For example, many factory and repair/maintenance jobs can be described in terms of a small number of psychomotor tasks. Researchers can identify these tasks via job analysis and then use the tasks as the basis for work sample selection tests. Unfortunately, there are often factors which make it impractical or even unwise to use such work sample tests as predictors.

First, as research by Adams (1953, 1957), Fleishman (1960; Fleishman & Hempel, 1954b, 1955; Parker & Fleishman, 1960), Hinrichs (1970), and others has demonstrated, the psychomotor abilities accounting for variance in task performance change as a task is practiced and learned. This implies that

the psychomotor tests which are valid as predictors of initial learning rate may have little or no relationship to a subject's asymptotic level of task performance. Indeed, if the abilities involved in initial task learning are totally different from the abilities involved in task performance after the task has been practiced for many trials, then work sample tests based on that task may be completely invalid as predictors of ultimate task proficiency.

Second, generally speaking, it is quite expensive to develop work sample psychomotor tests. This is due in part to the fact that many of the jobs requiring high levels of psychomotor abilities involve the use of extremely expensive equipment and machinery (e.g., pilots, truck drivers, and tank gunners). If the test setting is to correspond closely to the job setting, it may be necessary to develop very costly job simulators. Some researchers have circumvented this problem by sacrificing situational and/or task fidelity in their tests. For example, the Rudder Control apparatus developed by the AAF research team (Melton, 1947; see Appendix A) bears only a moderate resemblance to the cockpit of an airplane. Since the motor movements required to operate the apparatus are very similar to those required to fly an airplane, the test has good content validity (cf. Anastasi, 1976). Nevertheless, the use of a relatively inexpensive facsimile like the Rudder Control apparatus makes the test as much a "sign" as a "sample" of job performance. While operation of the Rudder Control apparatus certainly requires a complex integration of perceptual and psychomotor abilities, it does not even begin to compare with the perceptual-psychomotor demands of flying an airplane.

A third factor which decreases the utility of work sample tests is that most work sample tests are of rather limited generality. As noted above, the behaviors comprising the performance domain of any job typically require a complex blend of abilities which are almost never duplicated in any other job. This may not present test development problems if one is only attempting to predict job proficiency for a handful of closely related jobs which are each comprised of just three or four key tasks. For the current research effort, however, the Army is attempting to predict job proficiency for 19 different *job families*. Given our situation, it makes more sense to identify a finite set of personal constructs underlying performance on the various jobs and to develop "signs" tests to measure these constructs and to predict training performance or job proficiency.

Reliance on a construct-oriented test development strategy does not preclude consideration of content validity in the development of these tests. To the extent that it is possible to introduce tasks or elements of the job environment which are common across job families into the personal construct measures, this should be done. According to most experts, increasing the similarity of the tests and the jobs should increase the criterion-related validity of the tests. It should also increase the face validity of the tests, which may have a favorable impact upon subjects' motivation to perform well on the tests.

Finally, it should be noted that there is little evidence to support the contention that complex work sample tests are more valid predictors than "single-construct" measures. The validity data in Table 29 show that the median validity of complex psychomotor predictors across all job types and criteria was only .11 ($N=112$ validity coefficients). By comparison,

the median validity for all psychomotor abilities across all job types and criteria was .18 ($N=2,206$). Certainly the 112 validity coefficients for complex psychomotor predictors tabulated in Table 29 represent a less than exhaustive sample of the validity evidence for work sample tests. Nevertheless, given the high test development costs for most work sample tests, it would seem that much more evidence of the predictive validity of these tests would need to be accumulated before the expense of developing these measures can be justified.

In conclusion, it appears that there may be nothing to gain from investing a substantial amount of effort in the development of work sample tests. Work sample tests tend to be expensive to develop and there is no evidence that they are more valid than tests which are designed to measure ability constructs. All this suggests that a construct-oriented test development strategy, based on Fleishman's psychomotor ability taxonomy, would be the best strategy for this research program.

SECTION VI

SUMMARY AND RECOMMENDATIONS

Summary of Major Findings

Several taxonomies have been suggested for describing the psychomotor ability domain. By far the most popular taxonomy is that based on the factor analytic research of Edwin Fleishman and his associates (Fleishman, 1967, 1972). This review represents an effort to explicate and evaluate the predictive validity and practical utility of nine of the abilities identified by Fleishman: multilimb coordination, control precision, rate control, finger dexterity, manual dexterity, wrist-finger speed, aiming, arm-hand steadiness, and speed of arm movement.

Fleishman's taxonomy was selected for evaluation primarily because the abilities in this taxonomy seemed particularly relevant to the criteria of interest in this selection and classification research project. That is, the criteria in this study represent rather complex behaviors (e.g., Repair Mechanical Systems). It did not seem likely that extremely basic perceptual-motor processes from an information processing model of motor skills learning would be very powerful predictors of these criteria. On the other hand, it also seemed unlikely that a single broad psychomotor coordination dimension could adequately capture and predict the varied motor skills represented in the different criteria. The ability taxonomy suggested by Fleishman was therefore judged to be the most appropriate for this review.

One of the major problems with Fleishman's taxonomy is the lack of construct explication for the nine abilities. Indeed, this would have been a problem with any of the psychomotor ability taxonomies reviewed for this paper. Virtually no research has been devoted to identifying and developing a nomological net for any psychomotor construct. In one of the few scholarly attempts to explicate the psychomotor ability domain, Imhoff and Levine (1981) suggested that the abilities which control precision and speed of arm movement are most relevant to the types of motor skills under open-loop control (Keele, 1968), while multilimb coordination and rate control affect tasks under closed-loop control (Adams, 1971). Much work, however, remains to be done in this area.

To evaluate the probable criterion-related validity of these nine abilities, a comprehensive review of the literature was conducted. Over 2,300 validity coefficients were identified. These coefficients were categorized by predictor construct (i.e., ability), criterion, job type, and research setting (i.e., military vs. nonmilitary). An analysis of these validities revealed that performance in several job types was predictable from psychomotor abilities. For example: multilimb coordination and control precision demonstrated validity as predictors for professional, technical, and managerial jobs; manual dexterity and wrist-finger speed demonstrated validity for clerical jobs; multilimb coordination, manual dexterity, and wrist-finger speed demonstrated validity for protective service jobs; and finger dexterity, manual dexterity, and wrist-finger speed demonstrated validity for service jobs, mechanical and structural maintenance jobs, and industrial jobs.

An evaluation of the validity of psychomotor tests used in the military was not possible because psychomotor tests have not been widely employed by the Armed Forces. Almost 72% of the 547 psychomotor validity coefficients from military research which were located during the literature review were based on studies of pilots. No psychomotor validity coefficients were identified for many common Army jobs, including infantryman, administrative specialist, motor transport operator, and cannon crewman.

Overall, however, the validity picture for psychomotor tests is quite promising. The median of the 2,373 validity coefficients was .17, and the median validity of .20 for job proficiency criteria ($N=1,358$ validity coefficients) was actually greater than the median validity of .15 for training performance criteria ($N=603$ validity coefficients).

Since there are no data on the standard deviation of job performance in dollars for most Army jobs, and since the incremental validity of psychomotor tests (i.e., above and beyond the validity of current Army selection and classification procedures) for different Army jobs is unknown, it is currently impossible to assess the utility of psychomotor testing for the Army. It would appear, however, that psychomotor tests would have significant incremental validity.

As part of the literature review, 1,066 correlations between cognitive-perceptual and psychomotor tests were identified. These correlations were categorized by cognitive-perceptual ability and psychomotor ability. Median correlations were computed for each cognitive-perceptual/psychomotor ability combination. Analysis of these median correlations indicated that, with the possible exception of wrist-finger speed, there was only minimal overlap between the cognitive-perceptual and psychomotor abilities. Interestingly, two of the cognitive-perceptual abilities with the lowest correlations with psychomotor abilities were verbal comprehension and reasoning. These are also the two cognitive-perceptual abilities which are most highly correlated with general intelligence, g . Thus, there is apparently very little overlap between psychomotor abilities and g . These results suggest that any variance psychomotor tests share with criterion measures will be variance which cannot be tapped by any cognitive-perceptual ability test. Given the low correlations between psychomotor and cognitive-perceptual tests, any variance shared by psychomotor tests and criterion measures should result in significant increments in the overall validity of the test battery.

The stability of psychomotor tests has been a concern to some selection researchers. For example, Navy researchers have found that even the most stable psychomotor tests do not attain differential stability until subjects have had 30 minutes of practice per day for six days (Jones et al., 1981). It was shown, though, that differential stability (Bittner, 1979) is not the same as test-retest stability. When psychomotor test reliability coefficients were examined, it was found that the median test-retest reliability coefficient for intertest intervals ranging from immediate to 30 days was .78. While this may attenuate the validity coefficients for psychomotor tests somewhat, it is certainly not so serious a problem as to jeopardize the use of psychomotor tests in selection and classification research.

Other concerns with psychomotor tests have centered on the expense, unreliability, and transportability problems associated with psychomotor testing. Several times previously, different branches of the Armed Forces have decided against using psychomotor tests for selection and classification purposes because of the unreliability of the apparatus used to administer the tests. Mechanical breakdowns have occurred frequently. In addition, there have often been considerable disparities between different units of a given apparatus, resulting in significant differences in subjects' test scores as they are tested on different units. With the advent of microcomputer technology, most of the problems associated with cost, reliability, and transportability have been alleviated. This removes a major barrier to the widespread use of psychomotor tests.

A critical issue in all selection testing today is how members of different subgroups fare on the test. There have been very few studies of the impact of group differences on psychomotor test scores. Studies with the GATB revealed significant sex differences on two of the three psychomotor scales (U.S. Department of Labor, 1970). GATB research has also revealed significant differences between mean psychomotor test scores for different ethnic groups. Hispanics and Orientals tend to score highest on all three GATB scales, followed by American Indians and whites, with blacks scoring lowest (U.S. Department of Labor, 1970). The range between the mean scores of the highest and lowest scoring groups has averaged approximately one standard deviation. Results from AAF research conducted during World War II revealed only very small differences between mean apparatus test scores for blacks and whites (Melton, 1947). Tests for differential validity were not possible with the GATB because single-scale validity coefficients were not reported for different subgroups. In the AAF, there were no differences in test validities for men and women. In general, however, the psychomotor apparatus tests used in the ACB had virtually no validity for blacks and moderate to high validities for whites. This case represents one of the few examples of differential validity reported in the testing literature. It is based on only one sample of blacks, however, and the results are more than 40 years old. Thus, its implications for psychomotor test research today are uncertain. The results do suggest that statistical tests for differential validity should be conducted for any new psychomotor measures.

A Suggested Priority for Psychomotor Test Development Efforts

As noted in the introduction to this report, the goal throughout this literature review has been to identify a set of psychomotor abilities that are likely to be related to training performance and job proficiency criteria in the Army. Based in part on the recommendations resulting from this process, new psychomotor tests will be developed and incorporated into a new experimental Army selection and classification battery. The utility and validity of these new tests will then be evaluated over the next several years.

Several criteria were considered in making these recommendations. First, the validity of tests of each ability in previous studies was examined. Abilities which demonstrated moderate or high validities across several Army job types were given preference. This was largely a practical consideration. Because total testing time will be limited, and because all new recruits will receive the same test battery regardless of their stated

occupational preference, abilities were given priority if they were likely to be useful in predicting success in several Army job types.

A second criterion was to ensure that, insofar as possible, valid psychomotor tests were available for each job type. When only one or two abilities were valid predictors of performance for certain job types, those abilities were given extra consideration when making final recommendations. Third, the difficulty of developing and administering tests was evaluated for each ability. There was a strong desire to avoid any tests requiring individual administration and scoring. Therefore, preference was given to abilities which could be assessed easily via paper-and-pencil or computerized measures. Finally, the processes underlying each ability were considered. For example, an effort was made to ensure coverage of both open-loop control and closed-loop control phenomena in the experimental test battery. Other motor or perceptual requirements inherent in measures of each ability were also considered.

Based on these criteria, it is recommended that tests be developed according to the following prioritization:

First Priority

1. Multilimb coordination

Second Priority

2. Control precision
3. Manual dexterity
4. Rate control
5. Speed of arm movement

Third Priority

6. Wrist-finger speed
7. Finger dexterity

Fourth Priority

8. Arm-hand steadiness

Fifth Priority

9. Aiming

Multilimb Coordination. Multilimb coordination was given the highest priority for several reasons. First, tests of multilimb coordination have demonstrated high validity as predictors of training performance for both

protective service jobs and professional, technical, and managerial jobs. Second, multilimb coordination represents one of the psychomotor abilities related to performance of tasks under closed-loop control (Imhoff & Levine, 1981). Tests of multilimb coordination typically require subjects to monitor their performance closely and make continuous motor adjustments. Moreover, because tests of multilimb coordination require careful coordination of two limbs, multilimb coordination closely approximates what many people have in mind when they speak of a general psychomotor "klutziness" dimension. Thus, this ability is of very great theoretical interest. Finally, multilimb coordination lends itself readily to computerized assessment.

By definition, tests of multilimb coordination require the subject to use two or more limbs simultaneously. Not all tests which require the use of two limbs are measures of multilimb coordination, however. For example, the Complex Coordination Tests developed by Sanders et al. (1971) and Myers et al. (1982) require the subject to perform separate tracking tasks with the right and left hands. By contrast, many of the apparatus tests used during World War II (e.g., Two-Hand Coordination and Rudder Control) require subjects to use their hands and/or their hands and feet to accomplish but one task. The emphasis in the latter is on the *careful coordination* of two or more limbs, while the former emphasizes the performance of two tasks simultaneously. The former actually assesses time-sharing ability as much as multilimb coordination. While both types of tests are interesting, it is recommended that at least one "pure" measure of multilimb coordination be included in the experimental test battery.

In his factor analytic studies, Fleishman found that tests requiring the use of two hands and tests requiring the use of one hand and one foot both loaded on the same multilimb coordination factor. Thus, if only one multilimb coordination test is included in the experimental test battery, it should not matter which two limbs the subject is asked to use. Nevertheless, if it is possible to include two multilimb coordination tests, it would be best if one required the use of two hands and the other required the use of one hand and one foot.

Control Precision. Tests of control precision have demonstrated validity for predicting performance in protective service jobs and professional, technical, and managerial jobs, though the median validity for protective service jobs has only been moderate (median $r=.14$, $N=7$ validity coefficients). Tests of multilimb coordination, however, have demonstrated even higher validity for predicting performance in these two job types. Therefore, the validity of control precision tests was not the chief reason for recommending that these measures be included in the experimental test battery. Instead, control precision is interesting primarily because it appears to be related to performance of tasks controlled by motor programs (Imhoff & Levine, 1981). Thus, including control precision tests in the experimental test battery will ensure coverage of both open-loop control and closed-loop control processes. Like multilimb coordination, control precision lends itself readily to computerized testing.

The most popular test of control precision has been the Rotary Pursuit. For the Rotary Pursuit, the subject must track an object moving in a known direction at a constant rate of speed. Thus, one type of computerized control precision measure could involve a simple pursuit tracking

task in which the target moves in a known direction at a constant rate of speed. Another control precision test which has received some attention is Controls Adjustment. This test involves another type of simple "tracking" task. On each test trial, a stationary target stimulus is presented to the subject, who must manipulate a control stick as quickly as possible to move a pursuit stimulus so that its position matches that of the target stimulus. A computerized version of this test would serve as an excellent supplement to the more traditional control precision pursuit tracking task.

Manual Dexterity. Tests of manual dexterity have proven valid as predictors of performance for several job types, including clerical (median $r=.15$, $N=49$ validity coefficients), protective service (median $r=.15$, $N=5$ validity coefficients), service (median $r=.20$, $N=50$ validity coefficients), and mechanical and structural maintenance (median $r=.17$, $N=63$ validity coefficients). Thus, manual dexterity meets the most important criterion for inclusion in the experimental battery in that it has demonstrated validity for a number of different job types. Since little or no validity evidence was available for multilimb coordination and control precision for three of these job types, tests of manual dexterity would help ensure that valid psychomotor predictors were available for each Army job type. Manual dexterity also holds some theoretical interest because it probably ranks near multilimb coordination as an index of general "klutziness."

The major problem with existing manual dexterity tests is that they all require individual administration; no one has yet developed a manual dexterity test which can be administered in a group setting or via computer. The ever increasing number of peripheral devices available for microcomputers, including light pens, bit packs, and graphics tablets, may facilitate a creative solution to this problem.

Failing that, the best alternative would be to develop a test that combines elements of common finger and manual dexterity tests. Such a test might require subjects to perform some type of assembly and/or disassembly operation with blocks, similar to the assembly and disassembly tasks comprising the GATB Finger Dexterity Test. The only difference between this new test and the GATB Finger Dexterity Test would be that the blocks involved in the new test would be larger than the pegs involved in the GATB test. Thus, the subject would be required to make both fine, highly controlled manipulations and movements with the fingers (i.e., during assembly and disassembly) and more gross manipulations and movements with the hand (i.e., during lifting, turning, and placing). Since such a test would combine elements of manual and finger dexterity, it would alleviate the need to assess finger dexterity separately.

Rate Control. Rate control, like multilimb coordination, is involved in performance of tasks under closed-loop control. Rate control is probably a purer--or at least somewhat different--measure of closed-loop control processes than multilimb coordination is, since the motor response required in rate control tests is somewhat simpler than that required in multilimb coordination tests. This is because rate control tests do not require the subject to carefully coordinate the movements of two or more limbs. This difference between multilimb coordination and rate control makes rate control a theoretically interesting construct.

Unfortunately, there have been few investigations of the validity of

rate control tests. Moreover, the few validities which have been obtained have not been outstanding (median $r = .06$ across all job types and criteria, $N = 16$ validity coefficients). Thus, it is not readily apparent whether rate control would contribute to the predictive validity of the experimental test battery.

Nevertheless, a lack of validity research on rate control measures and a theoretical interest in this construct due to its association with closed-loop control processes suggest that rate control deserves additional attention and research. Therefore, it is recommended that rate control measures be included in the experimental test battery.

Rate control is usually assessed via tracking tests. The distinguishing feature of these tests is that the target stimulus is making continuous and unpredictable changes in speed and direction.

Two different types of tracking tests are commonly used. In a compensatory tracking test, subjects must keep a moving stimulus as close as possible to a fixed null point. In a pursuit tracking test, subjects control a pursuit stimulus which they attempt to keep centered about a target stimulus. It would be interesting to examine the validity of rate control tests involving both types of tracking tasks. If there is only time to administer one rate control test, however, the best procedure would be to ensure that both pursuit and compensatory tracking tests are included in the experimental test battery. Thus, the choice of a rate control test might depend on the types of tracking tests used in the measures of multi-limb coordination and control precision.

If time permits, a novel type of compensatory tracking test might also be included in the experimental test battery. In this test, a digital display instead of a moving target would be used as a stimulus. The subject's task would be to use a dial or sliding resistor to keep the display number as close to zero as possible. Such a procedure would appear to reduce the spatial requirements of the test. It would therefore be interesting to compare the validity and cognitive-perceptual correlates of such a test to the validity and cognitive-perceptual correlates of more traditional compensatory tracking tests.

Speed of Arm Movement. Two of the nine psychomotor abilities in our taxonomy were included by Imhoff and Levine (1981) in their basic movement speed and accuracy dimension. These abilities were control precision and speed of arm movement. According to Imhoff and Levine, both of these abilities are related to motor behavior controlled by open-loop processes. Yet, there is an important difference between control precision and speed of arm movement. In tests of control precision (e.g., Rotary Pursuit), the subject may be required to make very accurate movements. Scores on control precision tests are often error or accuracy measures (e.g., root-mean-square error, time on target). On the other hand, in tests of speed of arm movement (e.g., Two-Plate Tapping), the need for accurate movement is minimized. Targets in speed of arm movement tests are typically very large, so that virtually any movement in the direction of the target is likely to make contact with the target. The total emphasis on speed rather than accuracy makes speed of arm movement unique in the psychomotor ability domain. That is the chief reason for recommending that speed of arm movement be included in the experimental test battery.

Unfortunately, as Table 29 shows, there are almost no validity data for speed of arm movement. This makes it difficult to argue that a speed of arm movement test would contribute significantly to the prediction of performance in one or more job types. It also raises questions about the utility of developing a test specifically to assess this ability.

There is a way to keep any test development costs minimal, however. One measure considered for inclusion in the experimental computer battery is reaction time. In his recent research, Jensen has used a test apparatus which permits the assessment of both reaction time and movement time on each trial (e.g., Jensen, 1982). Subjects begin a trial with their index finger on a home button. When the stimulus is presented, subjects move their finger from the home button to a large target button. The time from the presentation of the stimulus until the release of the home button is the subject's reaction time. The time from the release of the home button until the target button is pressed is the movement time. Movement time is, of course, a measure of speed of arm movement.

In sum, even though there is no validity evidence for speed of arm movement, it is recommended that a measure of this ability be included in the experimental test battery. This is because speed of arm movement represents perhaps the only psychomotor ability where movement accuracy is not a factor. Moreover, speed of arm movement can be measured at the same time as reaction time. Thus, including a speed of arm movement measure in the experimental test battery will not increase total testing time for subjects and will result in only very minimal increases in test development costs.

Wrist-Finger Speed. Table 29 shows that wrist-finger speed has been a valid predictor of training performance and job proficiency for several job types, including clerical, protective service, and industrial. Nevertheless, there are at least three reasons for omitting tests of wrist-finger speed from the experimental test battery.

First, manual dexterity is also a valid predictor of performance for clerical, protective service, and industrial jobs. Thus, wrist-finger speed would not add to the number of job types for which valid psychomotor predictors are available in the experimental test battery.

Second, wrist-finger speed is not of compelling theoretical interest. Like control precision and speed of arm movement, wrist-finger speed is most closely related to movement controlled by open-loop processes. Thus, wrist-finger speed falls within Imhoff and Levine's (1981) basic movement speed and accuracy dimension. Within this dimension, speed of arm movement requires great speed and little accuracy and control precision requires approximately equal amounts of both speed and accuracy. Wrist-finger speed falls somewhere between the two, requiring somewhat more speed than accuracy. Wrist-finger speed would therefore probably not add greatly to our understanding of this dimension. Moreover, the movements involved in wrist-finger speed are quite simple. The movements do not require a great deal of precision, coordination, or timing.

Third, most tests of wrist-finger speed in current use must be scored by hand. While it might be possible to adapt wrist-finger speed tests to

the computer using a light pen, the technology is quite new and may not be sufficiently reliable to permit accurate assessment.

Because of these three problems and because total testing time is limited, it is recommended that no tests of wrist-finger speed be included in the experimental test battery.

Finger Dexterity. Table 29 shows that finger dexterity has also been a valid predictor of performance for several job types. Nevertheless, finger dexterity tests would contribute little to the experimental test battery. For example, finger dexterity tests would not add to the number of job types for which valid psychomotor predictors were available. In addition, finger dexterity tests would create even more serious test administration problems than wrist-finger speed tests because finger dexterity tests would have to be individually administered as well as hand scored. Thus, it is recommended that no finger dexterity tests be included in the experimental test battery.

Previously, it was noted that many finger dexterity tests require subjects to assemble or disassemble objects. It was suggested that assembly and disassembly tests could be incorporated into a manual dexterity test. In this manner, some of the individual differences associated with finger dexterity might be captured in the manual dexterity test. This would alleviate the need to administer separate tests of these two abilities. Given the expenses associated with the administration of dexterity tests, this appears to be the only practical solution to the finger dexterity testing problem.

Arm-Hand Steadiness. While the focus in measures of speed of arm movement is on movement speed, the focus in measures of arm-hand steadiness is on movement accuracy. In one of the two most common types of arm-hand steadiness tests, a subject's task is to trace through a narrow maze with a stylus without touching the sides of the maze. Arm-hand steadiness tests are generally not timed. Subjects can attempt to minimize error scores without worrying about time constraints. For this reason, arm-hand steadiness represents a rather interesting ability construct.

In the past, the predictive validity of arm-hand steadiness tests has received only limited attention. The results have generally been disappointing. Table 29 shows that validity data are available for just two job types, and median validities are less than .10 for both. Airplane pilot is perhaps the only job for which the validity of arm-hand steadiness has been investigated extensively. Since "smoothness of control movements" was identified as one of the key reasons for success in pilot training during World War II (see Table 4), it would seem reasonable that arm-hand steadiness tests should have at least moderate validity as predictors of training performance for pilots. Table 9 shows that this is not the case; the median validity of arm-hand steadiness tests for predicting attrition from pilot training was only .10 ($N=16$ validity coefficients). These results suggest that arm-hand steadiness would not contribute to the overall validity of the experimental test battery.

Another problem with arm-hand steadiness tests is that they would be expensive to administer. All of the arm-hand steadiness tests identified in the literature reviewed for this report were apparatus tests requiring

individual administration. Although it might be possible to develop a computerized arm-hand steadiness test, such a test would probably require some expensive hardware peripherals (e.g., a graphics tablet). The validity evidence suggests that this would not be a cost-effective investment in hardware.

Primarily because past research suggests that the validity of arm-hand steadiness tests would be negligible, it is recommended that no tests of arm-hand steadiness be included in the experimental test battery.

Aiming. Aiming is another psychomotor ability which has received very little attention in validity research. The very name of the ability, "aiming," is intriguing, though. The name suggests that this ability might be relevant to performance in a wide variety of Army combat MOS.

Closer inspection of the tests used to measure this ability, however, indicates that the ability may be misnamed--or, at the very least, that the name might be easily misinterpreted. A subject's task in an aiming test is to place a pencil dot inside a small circle. In some tests, the circles are numbered (e.g., from one to 100), and the subject must dot the circles in order. Regardless of the specific format and instructions of the tests, the subject's basic task does not change. The subject must still place pencil dots in small circles.

In an attempt to describe the ability underlying performance on these tests, Fleishman settled upon the name "aiming." Often he expanded on this name, referring to the ability as "aiming (eye-hand coordination)" in many of his factor analytic studies. Yet, there is no construct validity evidence to relate the ability to performance on any tasks requiring either aiming (e.g., aiming a gun) or eye-hand coordination (e.g., hitting a baseball or catching a football). Nor does there seem to be any reason to expect that the ability would generalize to any task other than aiming a pencil at a small circle. Indeed, that is probably a more appropriate name for this ability--"aiming a pencil at a small circle."

Because of the lack of either predictive or construct validity evidence, it is recommended that no measure of aiming be included in the experimental test battery.

Job Content and the Selection of Psychomotor Tests

As noted previously, the recommendations above are based primarily on the validity evidence presented in Tables 8-29. The ease and expense of administering and scoring measures of each ability were also considered, though these considerations were accorded somewhat less weight in deriving recommendations. The final choice of psychomotor abilities for the experimental test battery will require consideration of one additional source of information: the content of Army jobs.

There are several potential sources for this job content information. For example, as part of this selection and classification research project, training and job performance dimensions are being identified for a number of key Army MOS. Preliminary descriptions of many of these dimensions are currently available. In addition, there have been several opportunities for members of the test development research team to conduct job observa-

tions and to examine training simulators. Since these job observations have focused on MOS which require subjects to perform many skilled perceptual and motor tasks, this job content information should be especially relevant to our efforts to identify a final set of psychomotor abilities for the experimental test battery.

Job content information gathered from these sources might suggest that wrist-finger speed, finger dexterity, arm-hand steadiness, and/or aiming are vital to the performance of many Army jobs and should be measured in the experimental test battery. Conversely, the job content information might reveal that multilimb coordination, control precision, manual dexterity, rate control, and/or speed of arm movement are totally irrelevant to performance in the Army and need not be included in the experimental test battery. These issues, although beyond the scope of this report, must be considered to ensure that the appropriate psychomotor abilities are represented in the experimental test battery.

REFERENCES

- Ackerman, P. L., Schneider, W., & Wickens, C. D. (1982). *Individual differences in time-sharing ability: A critical review and analysis* (Report #8102). Champaign, IL: University of Illinois, Human Attention Research Laboratory.
- Adams, J. A. (1953). *The prediction of performance at advanced stages of training on a complex psychomotor task* (Research Bulletin 53-49). Lackland Air Force Base, TX: U.S. Air Force Human Resources Research Center.
- Adams, J. A. (1954). The effects of pacing on the learning of a psychomotor response. *Journal of Experimental Psychology*, 47, 101-105.
- Adams, J. A. (1956). *An evaluation of test items measuring motor abilities* (AFPTRC-TN-56-55). Lackland Air Force Base, TX: U.S. Air Force Personnel Training and Research Center.
- Adams, J. A. (1957). The relationship between certain measures of ability and the acquisition of a psychomotor criterion response. *Journal of General Psychology*, 56, 121-134.
- Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, 3, 111-149.
- Adams, J. A. (1976). Issues for a closed-loop theory of motor learning. In G. E. Stelmach (Ed.), *Motor control: Issues and trends*. New York: Academic Press.
- Adams, J. A., Goetz, E. T., & Marshall, P. H. (1972). Response feedback and motor learning. *Journal of Experimental Psychology*, 92, 391-397.
- Allen, T. W., Rose, A. M., & Kramer, L. J. (1978). *An information processing approach to performance assessment: III. An elaboration and refinement of an information processing battery* (AIR-58500-TR). Washington, DC: American Institutes for Research.
- Ammons, R. B. (1950). Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *Journal of Experimental Psychology*, 40, 777-787.
- Ammons, R. B., Ammons, C. H., & Morgan, R. L. (1958). Transfer of skill and decremented factors along the speed dimension in rotary pursuit. *Perceptual and Motor Skills*, 11, 43.
- Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: MacMillan.
- Anastasi, A. (1983). Evolving trait concepts. *American Psychologist*, 38, 175-184.
- Biers, D. W., & Sauer, D. W. (1982). *Job sample tests as predictors of M1 gunnery performance*. Dayton, OH: Systems Research Laboratories.

- Bilodeau, E. A., & Bilodeau, I. M. (1958). Variations of temporal intervals among critical events in five studies of knowledge of results. *Journal of Experimental Psychology*, 55, 603-612.
- Bilodeau, I. M. (1956). Accuracy of a simple positioning response with variation in the number of trials by which knowledge of results is delayed. *American Journal of Psychology*, 69, 434-437.
- Bittner, A. C. (1979). Statistical tests for differential stability. *Proceedings of the 23rd annual meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society.
- Bittner, A. C., Lundy, N. C., Kennedy, R. S., & Harbeson, M. M. (1982). Performance evaluation tests for environmental research (PETER): Spoke tasks. *Perceptual and Motor Skills*, 54, 1319-1331.
- Braune, R. J., & Wickens, C. D. (1983). The functional age profile: An objective decision criterion for the assessment of pilot performance. *Proceedings of the 2nd Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, Department of Aviation.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.
- Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, 37, 64-76.
- Brogden, H. E. (1949). When testing pays off. *Personnel Psychology*, 2, 171-183.
- Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic image. *Psychological Review*, 4, 27-53.
- Bullock, R. J., & Svyantek, D. J. (1985). Analyzing meta-analysis: Potential problems, an unsuccessful replication, and evaluation criteria. *Journal of Applied Psychology*, 70, 108-115.
- Campbell, C. H., & Black, B. A. (1982). *Predicting trainability of M1 crewmen* (FR-MTRD-82-7). Ft. Knox, KY: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, NJ: Erlbaum.
- Carter, R. C., Kennedy, R. S., & Bittner, A. C. (1981). Grammatical reasoning: A stable performance yardstick. *Human Factors*, 23, 587-592.
- Cascio, W. F. (1982). *Applied psychology in personnel management* (2nd ed.). Reston, VA: Prentice-Hall.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.

- Croll, P. R., Mullins, C. J., & Weeks, J. L. (1973). *Validation of the Cross-Cultural Aircrew Aptitude Battery on a Vietnamese pilot trainee sample* (AFHRL-TR-73-30). Lackland Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper and Row.
- Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, IL: University of Illinois Press.
- Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, 2, 153-166.
- Damos, D. L. (1978). Residual attention as a predictor of pilot performance. *Human Factors*, 20, 435-440.
- Damos, D. L., Kennedy, R. S., & Bittner, A. C. (1980). Development of a performance evaluation test for environmental research (PETER): Critical tracking task. *Proceedings of the 50th Annual Scientific Meeting of the Aerospace Medical Association*, 33-34.
- Davis, F. B. (1947). *The AAF Qualifying Examination*. Washington, DC: U. S. Government Printing Office.
- Delcomyn, F. (1980). Neural basis of rhythmic behavior in animals. *Science*, 210, 492-498.
- de Wet, D. R. (1959). Co-ordination and floating effect. *Journal of the National Institute of Personnel Research*, 8, 28-38.
- de Wet, D. R. (1960a). A portable hand-foot reaction test. *Journal of the National Institute of Personnel Research*, 8, 106-116.
- de Wet, D. R. (1960b). An improved steadiness apparatus and its validity for air-pilot selection. *Journal of the National Institute of Personnel Research*, 8, 122-136.
- de Wet, D. R. (1962). A variable co-ordination test and its potentiality as a gauge of aptitude for airmanship. *Psychologia Africana*, 9, 86-99.
- Dobbins, D. A., Martinek, H., Anderson, A. A., & Rosenberg, N. (1961). *Inter-and intra-apparatus variance of psychomotor test scores of Army helicopter pilot trainees* (AGO-HFRB-RM-61-5). Washington, DC: U.S. Army TAG Research and Development Command, Human Factors Research Branch.
- Dudek, F. J. (1948). The dependence of factorial composition of aptitude tests upon population differences among pilot trainees, I: The isolation of factors. *Educational and Psychological Measurement*, 8, 613-633.

- Dudek, F. J. (1949). The dependence of factorial composition of aptitude tests upon population differences among pilot trainees, II: The factorial composition of tests and criterion variables. *Educational and Psychological Measurement*, 9, 95-104.
- Dunnette, M. D., Rosse, R. L., Houston, J. S., Hough, L. M., Toquam, J., Lammlein, S., King, K. W., Bosshardt, M. J., & Keyes, M. (1982). *Development and validation of an industry-wide electric power plant operator selection system* (Technical Report 72). Minneapolis: Personnel Decisions Research Institute.
- Eaton, N. K., Johnson, J., & Black, B. A. (1980). *Job samples as tank gunnery performance predictors* (Technical Report 473). Ft. Knox, KY: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Eggenberger, J. C. (1976). *Pilot selection research system* (Report 76-7). Toronto: Canadian Forces Applied Personnel Research Unit.
- Equal Employment Opportunity Coordinating Council (1978). Uniform guidelines on employee selection procedures (1978). *Federal Register*, 43, 38290-38315.
- Eysenck, H. J. (1953). The logical basis of factor analysis. *American Psychologist*, 8, 105-114.
- Farr, J. L., O'Leary, B. S., Pfeiffer, C. M., Goldstein, I. L., & Bartlett, C. J. (1971). *Ethnic group membership as a moderator in the prediction of job performance: An examination of some less traditional predictors* (AIR-753-9/71-TR-2). Washington, DC: American Institutes for Research.
- Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, 48, 483-492.
- Fitts, P. M., & Posner, M. J. (1967). *Human performance*. Belmont, CA: Brooks/Cole.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Fleishman, E. A. (1953). Testing for psychomotor abilities by means of apparatus tests. *Psychological Bulletin*, 50, 241-261.
- Fleishman, E. A. (1954a). Dimensional analysis of psychomotor abilities. *Journal of Experimental Psychology*, 48, 437-454.
- Fleishman, E. A. (1954b). *Evaluations of psychomotor tests for pilot selection: The Direction Control and Compensatory Balance Tests* (AFPTRC-TR-54-131). Lackland Air Force Base, TX: U.S. Air Force Personnel and Training Research Center.

- Fleishman, E. A. (1957). A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *Journal of Applied Psychology*, 41, 263-272.
- Fleishman, E. A. (1958). Dimensional analysis of movement reactions. *Journal of Experimental Psychology*, 55, 438-453.
- Fleishman, E. A. (1960). Abilities at different stages of practice in rotary pursuit performance. *Journal of Experimental Psychology*, 60, 162-171.
- Fleishman, E. A. (1965). The prediction of total task performance from prior practice on task components. *Human Factors*, 7, 18-27.
- Fleishman, E. A. (1966). Human abilities and the acquisition of skill. In E. A. Bilodeau (Ed.), *Acquisition of skill*. New York: Academic Press.
- Fleishman, E. A. (1967). Performance assessment based on an empirically derived task taxonomy. *Human Factors*, 9, 349-366.
- Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist*, 27, 1017-1032.
- Fleishman, E. A. (1975). Toward a taxonomy of human performance. *American Psychologist*, 30, 1127-1149.
- Fleishman, E. A., & Ellison, G. D. (1962). A factor analysis of fine manipulative tests. *Journal of Applied Psychology*, 46, 96-105.
- Fleishman, E. A., & Ellison, G. D. (1969). Prediction of transfer and other learning phenomena from ability and personality measures. *Journal of Educational Psychology*, 60, 300-314.
- Fleishman, E. A., & Fruchter, B. (1960). Factor structure and predictability of successive stages of learning Morse Code. *Journal of Applied Psychology*, 44, 96-101.
- Fleishman, E. A., & Fruchter, B. (1965). Component and total task relations at different stages of learning a complex tracking task. *Perceptual and Motor Skills*, 20, 1305-1311.
- Fleishman, E. A., & Hempel, W. E. (1954a). A factor analysis of dexterity tests. *Personnel Psychology*, 7, 15-32.
- Fleishman, E. A., & Hempel, W. E. (1954b). Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 19, 239-252.
- Fleishman, E. A., & Hempel, W. E. (1955). The relation between abilities and improvement with practice in a visual discrimination reaction task. *Journal of Experimental Psychology*, 49, 301-312.

- Fleishman, E. A., & Hempel, W. E. (1956). Factorial analysis of complex psychomotor performance and related skills. *Journal of Applied Psychology*, 40, 96-104.
- Fleishman, E. A., & Ornstein, G. N. (1960). An analysis of pilot flying performance in terms of component abilities. *Journal of Applied Psychology*, 44, 146-155.
- Fleishman, E. A., & Parker, J. F. (1962). Factors in the retention and relearning of perceptual-motor skill. *Journal of Experimental Psychology*, 64, 215-226.
- Fleishman, E. A., & Rich, S. (1963). Role of kinesthetic and spatial-visual abilities in perceptual-motor learning. *Journal of Experimental Psychology*, 66, 6-11.
- Fowler, B. (1981). The Aircraft Landing Test: An information processing approach to pilot selection. *Human Factors*, 23, 129-137.
- Fozard, J. L., & Nuttall, R. L. (1971). General Aptitude Test Battery scores for men differing in age and socioeconomic status. *Journal of Applied Psychology*, 55, 372-379.
- Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.
- Grant, D. L., & Bray, D. W. (1970). Validation of employment tests for telephone company installation and repair occupations. *Journal of Applied Psychology*, 54, 7-14.
- Guilford, J. P., & Lacey, J. E. (Eds.) (1947). *Printed classification tests, Parts I and II* (Army Air Forces Aviation Psychology Program Research Report No. 5). Washington, DC: U.S. Government Printing Office.
- Guion, R. (1977). Content validity--the source of my discontent. *Applied Psychological Measurement*, 1, 1-10.
- Guion, R. (1978). "Content validity" in moderation. *Personnel Psychology*, 31, 205-213.
- Hall, J. F. (1982). *An invitation to learning and memory*. Boston: Allyn and Bacon.
- Halpin, A. W., & Winer, B. J. (1957). A factorial study of the leader behavior descriptions. In R. M. Stogdill and A. E. Coons (Eds.), *Leader behavior: Its description and measurement* (Research Monograph No. 88). Columbus, OH: Ohio State University, Bureau of Business Research.
- Hardesty, D., & Bevan, W. (1964). *Forms of orally-presented knowledge of results and serial reaction time* (Technical Report No. 18). San Diego: Office of Naval Research.

- Harrell, T. W. (1940). A factor analysis of mechanical ability tests. *Psychometrika*, 5, 17-33.
- Helmstadter, G. C. (1964). *Principles of psychological measurement*. New York: Appleton.
- Hempel, W. E., & Fleishman, E. A. (1955). A factor analysis of physical proficiency and manipulative skill. *Journal of Applied Psychology*, 39, 12-16.
- Hinrichs, J. R. (1970). Ability correlates in learning a psychomotor task. *Journal of Applied Psychology*, 54, 56-64.
- Hogan, J., & Eagan, M. (1983). *Potential psychomotor assessments for personnel decisions*. Unpublished manuscript, University of Tulsa.
- Hulin, C. L., & Alvares, K. M. (1971). *Three explanations of temporal changes in ability-skill relationships: Literature review and theoretical analysis* (AFHRL-TR-71-6). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.
- Humphreys, L. G. (1960). Investigations of the simplex. *Psychometrika*, 4, 313-323.
- Hunter, D. R. (1975). *Development of an enlisted psychomotor/perceptual test battery* (AFHRL-TR-77-28). Lackland Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- Hunter, D. R., & Thompson, N. A. (1978). *Pilot selection system development* (AFHRL-TR-78-33). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- Imhoff, D. L., & Levine, J. M. (1981). *Perceptual-motor and cognitive performance task battery for pilot selection* (AFHRL-TR-80-27). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.
- Inskeep, G. C. (1971). The use of psychomotor tests to select sewing machine operators--some negative findings. *Personnel Psychology*, 24, 707-714.
- Jensen, A. R. (1982). Reaction time and psychometric g. In H.J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.
- Jones, A. (1982). Providing the man in the back seat. *Proceedings of the 24th Annual Conference of the Military Testing Association*. San Antonio, TX: Military Testing Association.
- Jones, M. B. (1960). Molar correlational analysis. *U. S. Navy School of Aviation Medicine Monograph*, No. 4.
- Jones, M. B. (1962). Practice as a process of simplification. *Psychological Review*, 69, 274-294.
- Jones, M. B., Kennedy, R. S., & Bittner, A. C. (1981). A video game for performance testing. *American Journal of Psychology*, 94, 143-152.

- Joreskog, K. G. (1978). Structural analysis of covariance of correlation matrices. *Psychometrika*, 43, 443-477.
- Karlovich, R. S., & Graham, J. T. (1968). Auditorily paced keytapping performance during synchronous, delayed, and decreased visual feedback. *Perceptual and Motor Skills*, 26, 731-743.
- Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, 70, 387-403.
- Keele, S. W. (1973). *Attention and human performance*. Pacific Palisades, CA: Goodyear.
- Kennedy, R. S., & Bittner, A. C. (1977). The development of a Navy Performance Evaluation Test for Environmental Research (PETER). In L. T. Pope and D. Meister (Eds.), *Productivity enhancement: Personnel performance assessment in Navy systems*. San Diego: Naval Personnel Research and Development Center.
- Kennedy, R. S., & Bittner, A. C. (1980). The utility of commercially available television-computer games for assessing performance and other applications. *Proceedings of the 51st Annual Scientific Meeting of the Aerospace Medical Association*.
- Kennedy, R. S., Bittner, A. C., & Einbender, S. W. (1980). *Development of performance evaluation tests for environmental research (PETER): Trail making test*. New Orleans: Naval Biodynamics Laboratory.
- Kennedy, R. S., Bittner, A. C., Harbeson, M., & Jones, M. B. (1982). Television computer games: A "new look" in performance testing. *Aviation, Space and Environmental Medicine*, 53, 49-53.
- Kennedy, R. S., Bittner, A. C., & Jones, M. B. (1981). Videogame and conventional tracking. *Perceptual and Motor Skills*, 53, 310.
- Kennedy, R. S., Carter, R. C., & Bittner, A. C. (1980). A catalogue of performance tests for environmental research. *Proceedings of the 24th Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society.
- Kleiman, L. S., & Faley, R. H. (1978). Assessing content validity: Standards set by the Court. *Personnel Psychology*, 31, 701-713.
- Kohfeld, D. L. (1966). The prediction of perceptual-motor learning from independent verbal and motor measures. *Psychonomic Science*, 4, 413-414.
- Lackner, J. R. (1974). Proprioceptive facilitation of open-loop visual pointing. *Perceptual and Motor Skills*, 39, 263-265.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.

- Lashley, K. S., & Ball, J. (1929). Spinal conduction and kinesthetic sensitivity in the maze habit. *Journal of Comparative Psychology*, 9, 71-106.
- Laszlo, J. I. (1967). Training of fast tapping with reduction of kinesthetic, tactile, visual, and auditory sensations. *Quarterly Journal of Experimental Psychology*, 19, 344-349.
- Laszlo, J. I., & Manning, L. C. (1970). The role of motor programming, command, and standard in the central control of skilled movement. *Journal of Motor Behavior*, 2, 111-124.
- Lennenberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Lorge, I., & Thorndike, E. L. (1935). The influence of delay in the after-effect of a connection. *Journal of Experimental Psychology*, 18, 186-194.
- Mashburn, N. C. (1934). Mashburn automatic serial action apparatus for detecting flying aptitude. *Journal of Aviation Medicine*, 5, 155-160.
- Mathews, J. J., & Jensen, H. E. (1977). *Screening test battery for dental laboratory specialist course: Development and validation* (AFHRL-TR-77-53). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- McGrevey, D. F., & Valentine, L. D. (1974). *Validation of two aircrew psychomotor tests* (AFHRL-TR-74-4). Lackland Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- Melton, A. W. (Ed.) (1947). *Apparatus tests* (Army Air Forces Aviation Psychology Program Research Report No. 4). Washington, DC: U.S. Government Printing Office.
- Michael, W. B. (1949). Factor analyses of tests and criteria: A comparative study of two AAF pilot populations. *Psychological Monographs*, 63 (Whole No. 298).
- Miller, N. E. (Ed.) (1947). *Psychological research on pilot training* (Army Air Forces Aviation Psychology Program Research Report No. 8). Washington, DC: U.S. Government Printing Office.
- Mullins, C. J., Keeth, J. B., & Riederich, L. D. (1958). *Selection of foreign students for training in the United States Air Force* (AFHRL-TR-68-111). Lackland Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- Myers, D. C., Jennings, M. C., Schemmer, M., & Fleishman, E. A. (1982). *Validation of cognitive and perceptual-motor tests for assigning pilots to different missions: Development and administration of computer-interactive tests* (ARRO-3075-TR). Ft. Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Sciences.

- Naylor, J. C., & Briggs, G. E. (1963). *Long-term retention of learned skills: A review of the literature* (ASD Technical Report 61-390). Washington, DC: U.S. Department of Commerce.
- Newell, K. M. (1974). Knowledge of results and motor learning. *Journal of Motor Behavior*, 6, 235-244.
- North, R. A., & Griffin, G. R. (1977). *Aviator selection 1919-1977* (Special Report-77-2). Pensacola, FL: U.S. Naval Aerospace Medical Research Laboratory.
- Osborn, W. C., & Ford, J. P. (1976). *Research on methods of synthetic performance testing* (HumRRO-FR-CD(L)-76-1). Alexandria, VA: Human Resources Research Organization.
- Owens, J. M., Goodman, L. S., Pollack, J. G., & Braune, R. J. (1983). Testing for job proficiency in Naval aviation: Recent developments and new initiatives in performance-based screening. *Proceedings of the 25th Annual Conference of the Military Testing Association*. Pensacola, FL: Military Testing Association.
- Parker, J. F., & Fleishman, E. A. (1960). Ability factors and component performance measures as predictors of complex tracking behavior. *Psychological Monographs: General and Applied*, 74, 1-36.
- Parker, J. F., & Fleishman, E. A. (1961). Use of analytical information concerning task requirements to increase the effectiveness of skill training. *Journal of Applied Psychology*, 45, 295-302.
- Passey, G. E., & McLaurin, W. A. (1966). *Perceptual-psychomotor tests in aircrew selection: Historical review and advanced concepts* (PRL-TR-66-4). Lackland Air Force Base, TX: U.S. Air Force Personnel Research Laboratory.
- Payne, R. B., Rohles, F. H., & Cobb, B. B. (1952). *The pilot candidate selection program, IV: Test validities and intercorrelations* (Report No. 4). San Antonio, TX: U.S. Air Force School of Aviation Medicine and U. S. Navy.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical operations. *Journal of Applied Psychology*, 65, 373-406.
- Rim, Y. (1962). The predictive validity of seven manual dexterity tests. *Psychologia*, 5, 52-55.
- Roff, M. F. (1951). *Personnel and classification procedures: Spatial tests--a factorial analysis* (Project Report No. 21-29-002). San Antonio, TX: U.S. Air Force School of Aviation Medicine.
- Roff, M. F. (1953). A factorial study of tests in the perceptual area. *Psychometric Monographs*, No. 8.

- Sanders, J. H., Valentine, L. D., & McGrevey, D. F. (1971). *The development of equipment for psychomotor assessment* (AFHRL-TR-71-40). Lackland Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- Schmidt, R. A. (1975). *Motor skills*. New York: Harper and Row.
- Seales, D. M., Kennedy, R. S., & Bittner, A. C. (1980). Development of Performance Evaluation Tests for Environmental Research (PETER): Arithmetic computation. *Perceptual and Motor Skills*, 51, 1023-1031.
- Seashore, R. H., Buxton, C. E., & McCollom, I. N. (1940). Multiple factorial analysis of arm-hand precision test. *American Journal of Psychology*, 53, 251-259.
- Seibel, R. (1964). Levels of practice, learning curves, and system values for human performance on complex tasks. *Human Factors*, 6, 293-298.
- Shanthamani, V. S. (1978). Industrial use of dexterity tests. *Journal of Psychological Researches*, 23, 200-205.
- Singer, R. N. (1980). *Motor learning and human performance: An application to motor skills and movement behaviors* (3rd ed.). New York: Macmillan.
- Smode, A. F. (1958). Learning and performance in a tracking task under two levels of achievement information feedback. *Journal of Experimental Psychology*, 56, 297-304.
- Spearman, C. (1927). *The abilities of man*. London: Macmillan.
- Sternberg, R. J. (1981). Testing and cognitive psychology. *American Psychologist*, 36, 1181-1189.
- Sutter, E. L., Townsend, J. C., & Ornstein, G. N. (1954). *The light plane as a pre-primary selection and training device: II. Analysis of training data* (AFPTRC-TR-54-35). Lackland Air Force Base, TX: U.S. Air Force Personnel and Training Research Center.
- Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology*, 39, 212-222.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1. Chicago: University of Chicago Press.
- Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.
- U. S. Department of Labor, Employment and Training Administration (1977). *Dictionary of Occupational Titles* (4th ed.). Washington, DC: U.S. Government Printing Office.
- U. S. Department of Labor, Manpower Administration (1970). *Manual for the USTES General Aptitude Test Battery*. Washington, DC: U.S. Government Printing Office.

- Viteles, M. S. (1942). *An historical introduction to aviation psychology* (Report No. 4). Washington, DC: Civil Aeronautics Administration, Department of Research.
- Want, R. L. (1962). The validity of tests in the selection of air force pilots. *Australian Journal of Psychology*, 14, 133-139.
- Warner, W. L., Meeker, M., & Eels, K. (1960). *Social class in America: The evaluation of status*. New York: Harper.
- Weinberg, D. R., Guy, D. E., & Tupper, R. W. (1964). Variation of postfeedback interval in simple motor learning. *Journal of Experimental Psychology*, 67, 98-99.
- Weitzman, D. O. (1979). *Individual differences in geographic orientation and spatial memory representation*. Paper presented at the 87th Annual Convention of the American Psychological Association, New York.
- Welford, A. T. (1968). *Fundamentals of skill*. London: Methuen.
- Welford, A. T. (1976). *Skilled performance: Perceptual and motor skills*. Glenview, IL: Scott, Foresman.
- Wernimont, P. F., & Campbell, J. C. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372-376.
- Whitley, J. D. (1970). Effects of practice distribution on learning a fine motor task. *Research Quarterly*, 48, 576-583.
- Wickert, F. (1947). *Psychological research on problems of redistribution* (Army Air Forces Aviation Psychology Program Research Report No. 14). Washington, DC: U.S. Government Printing Office.
- Wiker, S. F., Kennedy, R. S., & Pepper, R. L. (1983). Development of Performance Evaluation Tests for Environmental Research (PETER): Navigational plotting. *Aviation, Space and Environmental Medicine*, 54, 144-149.
- Wilson, D. M. (1961). The central nervous control of flight in a locust. *Journal of Experimental Biology*, 48, 471-490.
- Wittenborn, J. R. (1945). Mechanical ability, its nature and measurement; II. Manual dexterity. *Educational and Psychological Measurement*, 5, 395-409.
- Zeidner, J., Martinek, H., & Anderson, A. A. (1961). *Evaluation of experimental predictors for selecting Army helicopter pilots--II* (Technical Research Note 101). Washington, DC: U.S. Army TAG Research and Development Command, Human Factors Research Branch.

APPENDIX A

Descriptions of Selected Psychomotor Tests

Table of Contents

<u>Psychomotor Test</u>	<u>Page</u>
Arm-Hand Steadiness Test	A-2
Complex Coordination Test (Melton, 1947)	A-4
Complex Coordination Test (Sanders et al., 1971)	A-8
GATB Finger Dexterity Scale	A-10
GATB Manual Dexterity Scale	A-13
GATB Motor Coordination Scale	A-16
Large Tapping Test	A-18
Minnesota Rate of Manipulation Test	A-19
Motor Judgment Test	A-21
O'Connor Tweezer and Finger Dexterity Test	A-23
Purdue Pegboard	A-25
Rate Control	A-27
Rotary Pursuit Test	A-29
Rudder Control Test	A-32
Santa Ana Finger Dexterity Test	A-35
Single-Dimension Pursuitmeter	A-38
Small Tapping Test	A-40
Steadiness Aiming Test	A-41
Trace Tapping II	A-43
Two-Hand Coordination Test (Melton, 1947)	A-44
Two-Hand Coordination Test (Sanders et al., 1971)	A-48
Two-Hand Pursuit Test	A-50
Two-Plate Tapping Test	A-53

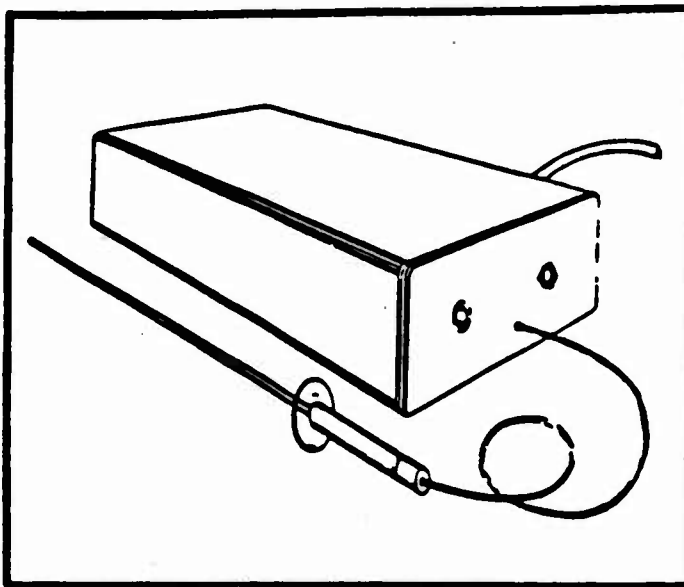
TEST TITLE: Arm-Hand Steadiness Test

**DESIGNED TO MEASURE
(CONSTRUCT):**

The steadiness of the arm and hand when the arm is extended at full length (Arm-Hand Steadiness)

**DESCRIPTION
OF TASK:**

The test apparatus consists of a metal rod or stylus which extends through a hole or aperture in a metal plate (see the figure below). The subject's task is to hold the stylus within the aperture without allowing the stylus to touch the sides or edge of the aperture.



**ADMINISTRATION
AND SCORING:**

The test consists of eight 30-second trials separated by 15-second rest periods. For the first trial, the subject uses his preferred hand. For the next six trials, the subject completes two trials with the nonpreferred hand followed by two trials with the preferred hand followed by two more trials with the nonpreferred hand. Finally, for the eighth trial, the subject again uses his preferred hand. The subject's score is either the number of contacts made with the sides or edge of the aperture or the total time in contact with the sides or edge of the aperture.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .76$ for time in contact score (Melton, 1947)

Mean correlation between three trials with the preferred hand, corrected for the full number of preferred hand trials (i.e., 3 trials). $N=310$ unclassified candidates for elementary pilot training.

$r_{xx} = .75$ for number of contacts score (Melton, 1947)

One-week test-retest reliability. $N=328$ unclassified candidates for elementary pilot training.

VALIDITY Median $r_{xy} = .06$ for total number of contacts score with graduation-elimination from pilot training (Melton, 1947)

$N=6,042$ Army Air Forces elementary pilot trainees in seven training classes.

Median $r_{xy} = .06$ for total number of contacts score with instructor ratings on six pilot training performance criteria (coordination, appropriate controls, feel of controls, smoothness of movement, progress in technique, probability of success in pilot training) (Melton, 1947)

$N=1,000$ Army Air Forces elementary pilot trainees.

Median $r_{xy} = -.03$ for total number of contacts score with graduation-elimination from navigator training (Melton, 1947)

$N=448$ Army Air Forces navigator trainees in three training classes.

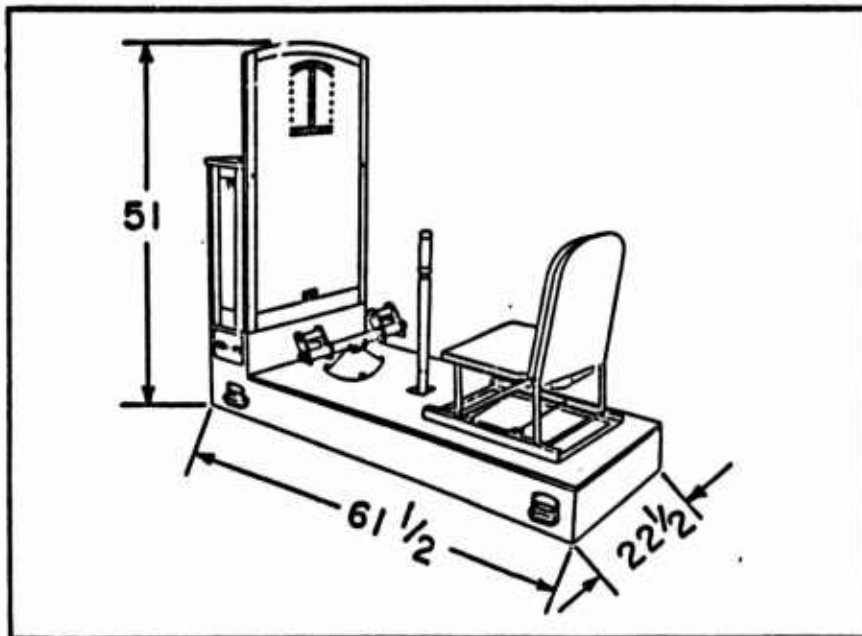
Median $r_{xy} = .06$ for total number of contacts score with with bombing accuracy score during bombardier training (Melton, 1947)

$N=331$ Army Air Forces bombardier trainees in three training classes.

TEST TITLE: Complex Coordination Test (Melton, 1947)

DESIGNED TO MEASURE
(CONSTRUCT): Ability to simultaneously coordinate the movements of the hands and the feet (Multilimb Coordination)

DESCRIPTION
OF TASK: Patterns of lights are presented whose positions are to be matched by appropriate adjustment of stick and rudder controls (see the figure below).



ADMINISTRATION
AND SCORING:

A correct response is accomplished only when both the hands and the feet have completed and maintained the appropriate adjustments, at which point a new pattern of lights to be matched is presented. Score is the number of completed matchings during two 2-minute test periods separated by a 30-second rest.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.89$ (Melton, 1947)

Odd-even split half reliability, corrected for the full length of the test (4 trials). $N=7,627$ unclassified candidates for pilot training.

$r_{xx}=.82$ (Fleishman, 1958)

Odd-even split half reliability, corrected for the full length of the test (2 trials). $N=204$ basic trainee airmen.

$r_{xx}=.87, .59, \text{ and } .83$ (Melton, 1947)

Immediate, one-week, and 28-day test-retest reliabilities. $N=415, 313, \text{ and } 692$ unclassified candidates for pilot training, respectively.

VALIDITY

Median $r_{xy}=.29$ with graduation-elimination from elementary pilot training (Melton, 1947)

$N=44,618$ Army Air Forces elementary pilot trainees in 24 training classes.

$r_{xy}=.35$ with graduation-elimination from elementary pilot training (Craeger, 1957)

$N=2,010$ Air Force elementary pilot trainees.

Median $r_{xy}=.23$ with graduation-elimination from single engine jet flying training (Leiman and Friedman, 1952)

$N=3,538$ Air Force advanced pilot trainees in two training classes.

Median $r_{xy}=.38$ with graduation-elimination from pilot training (Payne et al., 1952)

$N=1,345$ Navy pilot trainees in two training classes.

$r_{xy}=.31 \text{ and } .36$ with graduation-elimination from primary pilot training (Zaccaria and Cox, 1952)

$N=1,016$ aviation cadets and 547 student officers, respectively, in primary pilot training.

VALIDITY
(cont'd)

Median $r_{xy} = .43$ with graduation-elimination from primary pilot training (Fleishman, 1954b)

$N = 4,311$ Air Force pilot trainees in two training classes.

$r_{xy} = .23$ with graduation-elimination from pilot training (Mullins et al., 1968)

$N = 120$ foreign military pilot trainees receiving training from the U. S. Air Force.

$r_{xy} = .13$ with average check ride grade (Melton, 1947)

$N = 311$ Army Air Forces pilot trainees.

Median $r_{xy} = .21$ with instructor ratings on six pilot training performance criteria (coordination, appropriate controls, feel of controls, smoothness of movement, progress in technique, probability of success in pilot training) (Melton, 1947)

$N = 1,000$ Army Air Forces elementary pilot trainees.

$r_{xy} = .13$ with instructor ratings of flying proficiency (Melton, 1947)

$N = 562$ Army Air Forces advanced pilot trainees training on single-engine planes.

$r_{xy} = .13$ and $.07$ with instructor ratings of flying proficiency during the advanced and the transition stage of flying training, respectively

$N = 685$ Army Air Forces advanced pilot trainees training on four-engine planes.

Median $r_{xy} = .10$ with gunnery proficiency scores obtained during training (Melton, 1947)

$N = 1,716$ Army Air Forces advanced pilot trainees in four training classes.

VALIDITY
(cont'd)

Median $r_{xy} = .25$ with graduation-elimination from helicopter training (Zeidner et al., 1958)

N=249 Army helicopter pilot trainees in two training classes.

Median $r_{xy} = .20$ with graduation-elimination from navigator training (Melton, 1947)

N=1,752 Army Air Forces navigator trainees in two training classes.

Median $r_{xy} = .26$ with graduation-elimination from bombardier training (Melton, 1947)

N=4,333 Army Air Forces bombardier trainees in nine training classes.

$r_{xy} = .24$ with percent of "on target" bombs in combat (Melton, 1947)

N=32 Army Air Forces lead bombardiers.

$r_{xy} = .16$ with gunnery accuracy (Melton, 1947)

N=164 Army Air Forces trained B-29 remote-control turret gunners.

Median $r_{xy} = .18$ with composite flight and ground trainer grades (Melton, 1947)

N=450 Army Air Forces radar operator trainees in two training classes.

Median $r_{xy} = .06$ with several archival measures of driving performance (e.g., number of accidents, number of moving violations) (Farr et al., 1971)

N=299 taxi cab drivers.

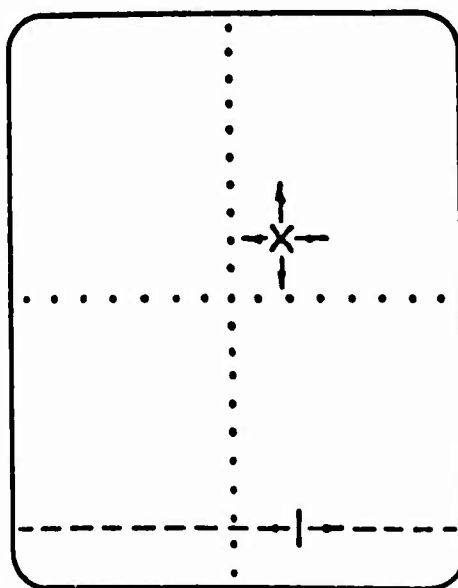
TEST TITLE: Complex Coordination Test (Sanders et al., 1971)

DESIGNED TO MEASURE
(CONSTRUCT):

The speed at which a subject makes complex hand and foot motor adjustments to a series of signal patterns (Multilimb Coordination)

DESCRIPTION
OF TASK:

This is a computer-administered test. The subject's task is to use joysticks and a foot-controlled rudder to maintain two stimuli in a null position (see the figure below). The movement of both stimuli is partially under the control of a computer program. This program causes the stimuli to make frequent, unpredictable changes in speed and direction. The first stimulus, which is X-shaped, can also be controlled via a joystick. The subject must attempt to keep this stimulus centered at the intersection of the row and column of dots. The direction and displacement of the joystick control the direction of movement and velocity, respectively, of the first stimulus. The second stimulus is a short vertical line which appears near the bottom of the display. Movement of this stimulus can be controlled by a rudder bar. The subject must use both feet to manipulate this rudder bar. The direction and displacement of the rudder bar control the direction and velocity, respectively, of the second stimulus. The second stimulus can only move along the horizontal dimension. The subject's task is to keep the second stimulus aligned along the vertical row of dots.



ADMINISTRATION
AND SCORING:

The test consists of five 1-minute trials. There is no rest period between trials; the test is continuous. The subject receives five error scores: (1) horizontal deviation of the first stimulus from the target point, called the X Axis score; (2) vertical deviation of the first stimulus from the target point, called the Y Axis score; (3) the square root of the sum of squares of the X Axis and Y Axis error scores (i.e., the Euclidean distance), called the Generated score; (4) the horizontal deviation of the second stimulus from the target point, called the Z Axis score; and (5) the number of times the subject allows the second stimulus to move off the screen, called the Reset score.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.92$ for the Generated score (Sanders et al., 1971)

Correlation between scores from minutes 4 and 5 of the test, corrected for length to represent the reliability of the sum of the two scores. $N=120$ Air Force officer trainees.

VALIDITY Median $r_{xy}=.19, .24, .20, .22,$ and $.13$ with graduation-elimination from pilot training for the X Axis score, Y Axis score, Generated score, Z axis score, and Reset score, respectively (McGrevey and Valentine, 1974)

$N=213$ Air Force officer undergraduate pilot trainees in two training classes.

$r_{xy}=.18$ with graduation-elimination from pilot training (Bory and Goodman, in preparation)

$N=294$ Navy undergraduate pilot trainees.

TEST TITLE: General Aptitude Test Battery (GATB) Finger Dexterity Scale

DESIGNED TO MEASURE
(CONSTRUCT): Finger Dexterity

DESCRIPTION
OF TASKS: Assemble Test. The board consists of 50 holes and two trays containing rivets and washers, respectively. The subject must pick up a rivet with the preferred hand and a washer with the non-preferred hand. The subject must then insert the rivet through one of the holes in the board and secure it with the washer.

Disassemble Test. The bottom test board consists of 50 rivets secured into holes with washers, and the top test board consists of 50 holes. The subject must remove the washer from the rivet, place the washer on a ring, remove the rivet from the hole, and place the rivet in an empty hole in the top test board.

ADMINISTRATION
AND SCORING: The subject's score is the number of parts assembled and disassembled in one and one-half minutes and in one minute, respectively.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.65$ and $.69$ (U. S. Department of Labor, 1970)
Three-month test-retest reliability for 605 male and 605 female high school seniors, respectively.

$r_{xx}=.76$, $.68$, and $.74$ (U. S. Department of Labor, 1970)
One-year, 2-year, and 3-year test-retest reliability for a sample of government employees. N unknown.

VALIDITY Median $r_{xy}=.12$ across educational, training, and job proficiency criteria for professional, technical, and managerial jobs (Droege, 1968; U. S. Department of Labor, 1970)

$N=105$ validity studies.

VALIDITY
(cont'd)

Median r_{xy} = .13 across educational, training, and job proficiency criteria for clerical jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 47 validity studies.

Median r_{xy} = .08 across educational, training, and job proficiency criteria for sales jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 5 validity studies.

Median r_{xy} = .10 across educational, training, and job proficiency criteria for protective service jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 5 validity studies.

Median r_{xy} = .14 across educational, training, and job proficiency criteria for service jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 47 validity studies.

Median r_{xy} = .20 across educational, training, and job proficiency criteria for mechanical and structural maintenance jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 59 validity studies.

Median r_{xy} = .16 across educational, training, and job proficiency criteria for electronics jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 10 validity studies.

Median r_{xy} = .25 across educational, training, and job proficiency criteria for industrial jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 257 validity studies.

VALIDITY
(cont'd)

Median $r_{xy} = .18$ across educational, training, and job proficiency criteria for miscellaneous jobs (e.g., farm worker, power plant operator) (Droege, 1968; U. S. Department of Labor, 1970)

$N=27$ validity studies.

TEST TITLE: General Aptitude Test Battery (GATB) Manual Dexterity Scale

DESIGNED TO MEASURE
(CONSTRUCT): Manual Dexterity

DESCRIPTION
OF TASKS:

Placing Test. The test apparatus consists of two boards, each containing 48 holes (four rows with 12 holes in each row). The holes on one of the boards are filled with pegs which fit snugly into the holes. The subject's task is to move all of the pegs from the holes of the first board to the holes of the second board as quickly as possible.

Turning Test. This test requires only one board. The holes of the board are all filled with pegs before the test begins. The subject's task is to remove pegs from the holes with one hand, turn the pegs over with the second hand, and re-insert the pegs into the holes with the second hand as quickly as possible.

ADMINISTRATION
AND SCORING:

The subject is given three 15-second placing trials and three 30-second turning trials. Trials are separated by 15-second rest periods. The subject's score is always the number of pegs removed from their holes, regardless of whether the pegs are placed or turned properly.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.73$ and $.72$ (U. S. Department of Labor, 1970)

Three-month test-retest reliability for 605 male and 605 female high school seniors, respectively.

$r_{xx}=.76$, $.72$, and $.78$ (U. S. Department of Labor, 1970)

One-year, 2-year, and 3-year test-retest reliability for a sample of government employees. N unknown.

VALIDITY Median $r_{xy}=.12$ across educational, training, and job proficiency criteria for professional, technical, and managerial jobs (Droege, 1968; U. S. Department of Labor, 1970)

$N=105$ validity studies.

VALIDITY
(cont'd)

Median $r_{xy} = .13$ across educational, training, and job proficiency criteria for clerical jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=47 validity studies.

Median $r_{xy} = .08$ across educational, training, and job proficiency criteria for sales jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=5 validity studies.

Median $r_{xy} = .15$ across educational, training, and job proficiency criteria for protective service jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=5 validity studies.

Median $r_{xy} = .18$ across educational, training, and job proficiency criteria for service jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=47 validity studies.

Median $r_{xy} = .20$ across educational, training, and job proficiency criteria for mechanical and structural maintenance jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=60 validity studies.

Median $r_{xy} = .10$ across educational, training, and job proficiency criteria for electronics jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=10 validity studies.

Median $r_{xy} = .25$ across educational, training, and job proficiency criteria for industrial jobs (Droege, 1968; U. S. Department of Labor, 1970)

N=257 validity studies.

VALIDITY
(cont'd)

Median $r_{xy} = .10$ across educational, training, and job proficiency criteria for miscellaneous jobs (e.g., farm worker, power plant operator) (Droege, 1968; U. S. Department of Labor, 1970)

$N=28$ validity studies.

TEST TITLE: General Aptitude Test Battery (GATB) Motor Coordination Scale

DESIGNED TO MEASURE (CONSTRUCT): Ability to coordinate eyes and hands or fingers rapidly and accurately in making precise movements with speed (Wrist-Finger Speed)

DESCRIPTION OF TASK: This test consists of a series of squares in which the subject is to make three pencil marks, working as rapidly as possible. The marks to be made are short lines, two vertical and the third a horizontal line beneath them.

ADMINISTRATION AND SCORING: The subject's score is the number of squares which are marked in 60 seconds.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .76$ and $.86$ (U. S. Department of Labor, 1970)
Three-month test-retest reliability for 605 male and 605 female high school seniors, respectively.

$r_{xx} = .86$, $.85$, and $.88$ (U. S. Department of Labor, 1970)
One-year, 2-year, and 3-year test-retest reliability for a sample of government employees. N unknown.

VALIDITY Median $r_{xy} = .12$ across educational, training, and job proficiency criteria for professional, technical, and managerial jobs (Droege, 1968; U. S. Department of Labor, 1970)

$N = 97$ validity studies.

Median $r_{xy} = .14$ across educational, training, and job proficiency criteria for clerical jobs (Droege, 1968; U. S. Department of Labor, 1970)

$N = 49$ validity studies.

VALIDITY
(cont'd)

Median r_{xy} = .14 across educational, training, and job proficiency criteria for sales jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 5 validity studies.

Median r_{xy} = .23 across educational, training, and job proficiency criteria for protective service jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 5 validity studies.

Median r_{xy} = .20 across educational, training, and job proficiency criteria for service jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 47 validity studies.

Median r_{xy} = .16 across educational, training, and job proficiency criteria for mechanical and structural maintenance jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 59 validity studies.

Median r_{xy} = .14 across educational, training, and job proficiency criteria for electronics jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 10 validity studies.

Median r_{xy} = .21 across educational, training, and job proficiency criteria for industrial jobs (Droege, 1968; U. S. Department of Labor, 1970)

N = 256 validity studies.

Median r_{xy} = .22 across educational, training, and job proficiency criteria for miscellaneous jobs (e.g., farm worker, power plant operator) (Droege, 1968; U. S. Department of Labor, 1970)

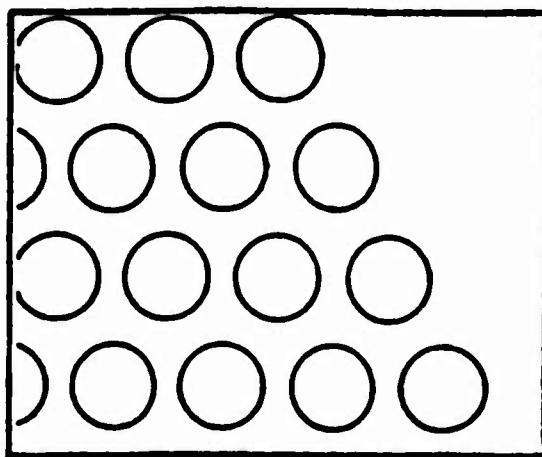
N = 28 validity studies.

TEST TITLE: Large Tapping Test

DESIGNED TO MEASURE
(CONSTRUCT): Wrist-Finger Speed

DESCRIPTION
OF TASK:

The subject is presented with a piece of paper containing six blocks. Each block in turn contains four rows of 10 large circles. The subject's task is to place three pencil marks or dots in each circle as rapidly as possible. (The figure below is similar, but not identical, to a portion of a block from this test.)



ADMINISTRATION
AND SCORING:

The subject's score is the number of circles marked with three dots at the end of two minutes.

PSYCHOMETRIC DATA:

RELIABILITY No data are available.

VALIDITY $r_{xy} = .05$ with graduation-elimination from pilot training
(Hunter and Thompson, 1978)

$N = 245$ Air Force pilot trainees.

$r_{xy} = .09$ with graduation-elimination from pilot training
(Croll, 1973)

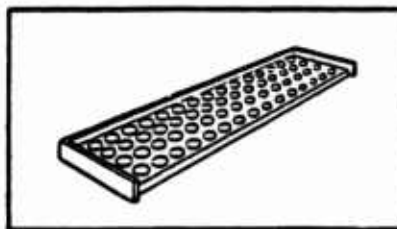
$N = 120$ South Vietnamese Air Force pilot trainees
receiving training from the U. S. Air Force.

TEST TITLE: Minnesota Rate of Manipulation Test

DESIGNED TO MEASURE
(CONSTRUCT): Ability to make gross arm-hand manipulations (Manual Dexterity)

DESCRIPTION
OF TASKS:

Placing Test. The test apparatus consists of two boards, each containing 60 wells (four rows with 15 wells in each row; see the figure below). The wells on one of the boards are filled with blocks which fit snugly into the wells. The subject's task is to move all of the blocks from the wells of the first board to the wells of the second board as quickly as possible.



Turning Test. This test requires only one board. At the beginning of the test, the wells of the board are filled with blocks. The subject's task is to remove the blocks from the wells with one hand, turn the blocks over with the second hand, and re-insert the blocks into the wells as quickly as possible with the second hand.

ADMINISTRATION
AND SCORING:

The subject is given five trials for each test. The first trial is a practice trial. The total time required for the subject to place or turn all 60 blocks on the remaining four trials is the subject's score on the test.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .87$ for placing and $.79$ for turning (Parker and Fleishman, 1960)

The method of estimation is unknown. Since the test consists of two placing trials and two turning trials, and since the reliability estimates appear to be based on a single test administration, it is likely that these reliabilities represent the correlation between the two trials corrected for total test length. $N=203$ freshman and sophomore U. S. Air Force ROTC students.

VALIDITY

Partial $r_{xy} = .30$ with grades in a shop mechanics course (Rim, 1962)

$N=224$ shop students in an Israeli high school. Predictor was the sum of placing, turning, and displacing test scores. The variance attributable to a paper-and-pencil intelligence test was partialled from both the predictor and the criterion.

$r_{xy} = .24$ with supervisory ratings of quickness in wrapping soap (Shanthamani, 1978)

$N=60$ soap packers in a factory in India. Predictor was the sum of the placing and turning scores.

$r_{xy} = .52$ with individual production records (Shanthamani, 1978)

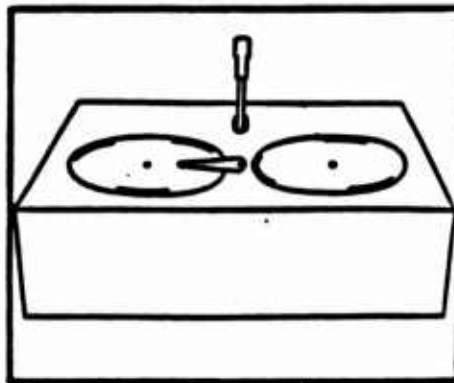
$N=15$ soap packers in a factory in India. Predictor was the sum of the placing and turning scores.

TEST TITLE: Motor Judgment Test

**DESIGNED TO MEASURE
(CONSTRUCT):** Rate Control

**DESCRIPTION
OF TASK:**

The subject is confronted by two adjacent disks rotating at a constant speed. Each disk has black and white sections on its perimeter (see the figure below). Between these disks is a pointer. The subject can control the speed of movement of this pointer with the control stick. A forward movement of the stick slows the pointer and a backward movement of the stick speeds the pointer. The subject cannot totally halt the rotation of the pointer and he cannot exert any control over the rotation of the two disks. The subject must manipulate the control stick so as to make as many rotations as possible without crossing the black areas on the rotating disks.



**ADMINISTRATION
AND SCORING:**

The subject's score is the ratio of the number of pointer revolutions to the number of errors (crossings of the black areas on the rotating disks) during four 1-minute trials. The trials are separated by 15-second rest periods.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .96$ and $.92$ (Melton, 1947)

Odd-even split half reliability for the errors score and revolutions score, respectively, corrected for the full length of the test (3 trials). $N=50$ qualified aircrew candidates.

$r_{xx} = .76$ (Fleishman, 1958)

Odd-even split half reliability for the ratio of errors to revolutions, corrected for the full length of the test (4 trials). $N=204$ basic trainee airmen.

VALIDITY Median $r_{xy} = -.02$ with several archival measures of driving performance (e.g., number of accidents, number of moving violations) (Farr et al., 1971)

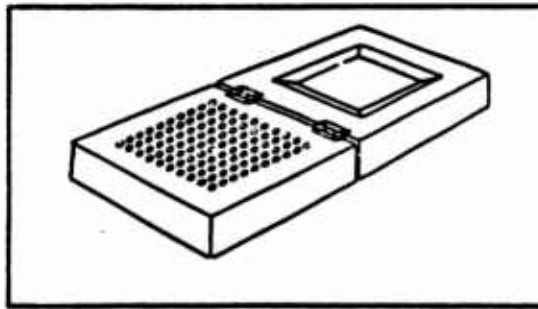
$N=301$ taxi cab drivers.

TEST TITLE: O'Connor Tweezer and Finger Dexterity Test

DESIGNED TO MEASURE
(CONSTRUCT): Finger Dexterity

DESCRIPTION
OF TASK:

The test board consists of one hundred 3/16-inch holes (see the figure below). The subject's task is to place three 1/16-inch solid brass pins in each hole as quickly as possible.



ADMINISTRATION
AND SCORING:

The subject is given three minutes to place the pins in the holes. The test score is the number of holes filled with three pins at the end of the three minutes.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .76$ (Parker and Fleishman, 1960)

Method of estimation is unknown. $N=203$ freshman and sophomore U. S. Air Force ROTC students.

VALIDITY Median $r_{xy} = .18$ with supervisory ratings of overall job performance (Laney, 1951)

$N=60$ gas appliance service workers from two gas utility companies.

Median $r_{xy} = .04$ with piece-rate pay (Inskeep, 1971)

$N=1,092$ female sewing machine operators in eight different plants.

VALIDITY
(cont'd)

Partial $r_{xy} = .33$ with grades in a shop mechanics course
(Rim, 1962)

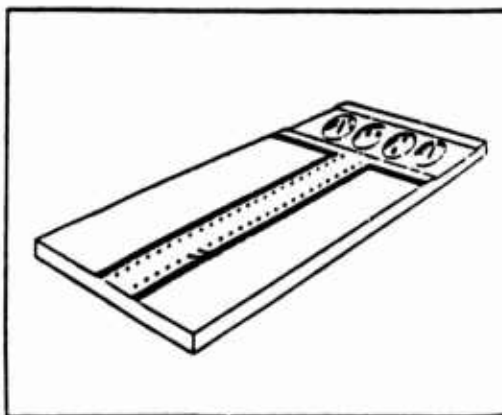
$N=224$ shop students in an Israeli high school. The variance attributable to a paper-and-pencil intelligence test was partialled from both the predictor and the criterion.

TEST TITLE: Purdue Pegboard

DESIGNED TO MEASURE
(CONSTRUCT): Finger Dexterity

DESCRIPTION
OF TASK:

The test apparatus consists of a board with 50 small holes (see the figure below). At the end of the board is a tray containing small dishes of pegs, collars, and washers.



Right hand score. The subject must pick up one peg at a time from the dish with his right hand and insert the peg into one of the holes in the board. The subject's score is the number of pegs inserted in one 30-second trial.

Left hand score. This is the same as the right hand score, except the subject must use his left hand to manipulate the pegs.

Both hands score. The subject must pick up two pegs at a time from the dish, one with his right hand and one with his left hand, and insert them into holes in the board. The score is the number of pegs inserted in one 30-second trial.

Assembly score. The subject must assemble peg-washer-collar combinations as quickly as possible. The subject's score is the number of combinations assembled in one 30-second trial.

Summation score. This is the sum of the four scores above.

ADMINISTRATION
AND SCORING:

See above.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .90$ for the Summation Score (Parker and Fleishman, 1960)

The authors state that this reliability is "based on (the) average reliability of the four subparts of this test and corrected for an increase in length by a factor of four" (Parker and Fleishman, 1960, p.3). The method of estimating the reliability of the four subparts is unknown. $N=203$ freshman and sophomore U. S. Air Force ROTC students.

VALIDITY Partial $r_{xy} = .70$ with grades in a shop mechanics course (Rim, 1962)

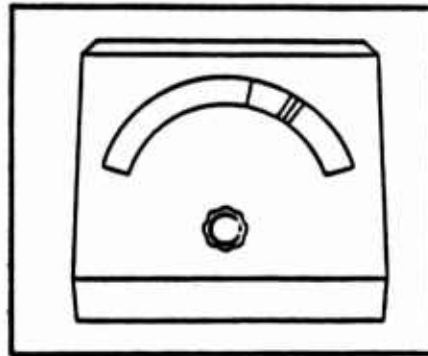
$N=224$ shop students in an Israeli high school. The variance attributable to a paper-and-pencil intelligence test was partialled from both the predictor and the criterion.

TEST TITLE: Rate Control

DESIGNED TO MEASURE
(CONSTRUCT): Rate Control

DESCRIPTION
OF TASK:

The test apparatus consists of a box containing a curved scale (see the figure below). A vertical target line moves back and forth across this scale. The line makes frequent, unpredictable changes in speed and direction. Centered below the scale is a rotary knob. This knob controls movement of a thin pointer line along the scale. The subject must attempt to keep this pointer line in coincidence with the target line as the target line moves about.



ADMINISTRATION
AND SCORING:

The subject's score on this test is the total time the target and pointer lines are in coincidence across eight one-minute trials.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .81$ (Melton, 1947)

Odd-even split-half reliability, corrected for the full length of the test (8 trials). $N = 381$ airborne radar students.

$r_{xx} = .69$ (Fleishman, 1958)

Odd-even split-half reliability, corrected for the full length of the test (4 trials). $N = 204$ basic trainee airmen.

VALIDITY

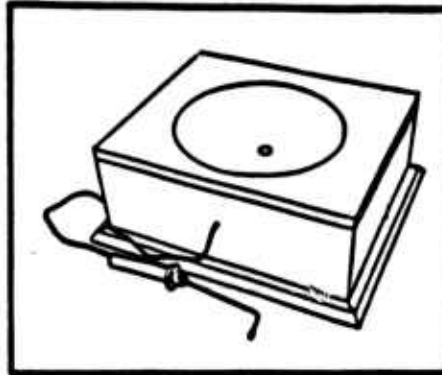
$r_{xy} = .02$ with graduation-elimination from radar operator training (Melton, 1947)

$N = 381$ Army Air Forces radar operator trainees.

TEST TITLE: Rotary Pursuit Test

DESIGNED TO MEASURE
(CONSTRUCT): Control Precision

DESCRIPTION
OF TASK: The subject must attempt to keep a stylus tip in contact with a small metallic target while the target is revolving near the edge of a phonograph-like turntable (see the figure below).



ADMINISTRATION
AND SCORING: The subject's score is the total time on target over five 20-second trials. The trials are separated by 10-second rest periods.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .98$ (Melton, 1947)

Odd-even split-half reliability, corrected for the full length of the test (20 trials). $N=301$ unclassified candidates for pilot training.

$r_{xx} = .81$ (Fleishman, 1958)

Odd-even split-half reliability, corrected for the full length of the test (5 trials). $N=204$ basic trainee airmen.

$r_{xx} = .88$ (Melton, 1947)

Immediate test-retest reliability. $N=398$ unclassified candidates for pilot training.

VALIDITY

$r_{xy} = .03$ with graduation-elimination from pilot training (Mullins et al., 1968)

$N = 120$ foreign military pilot trainees receiving training from the U. S. Air Force.

Median $r_{xy} = .22$ with graduation-elimination from elementary pilot training (Melton, 1947)

$N = 12,884$ Army Air Forces elementary pilot trainees in 10 training classes.

Median $r_{xy} = .27$ with graduation-elimination from primary pilot training (Fleishman, 1954b)

$N = 4,311$ Air Force pilot trainees in two training classes.

$r_{xy} = .01$ with instructor ratings of flying proficiency (Melton, 1947)

$N = 562$ Army Air Forces advanced pilot trainees.

Median $r_{xy} = .02$ with gunnery proficiency scores obtained during training (Melton, 1947)

$N = 1,193$ Army Air Forces advanced pilot trainees in three training classes.

Median $r_{xy} = .22$ with graduation-elimination from helicopter pilot training (Zeidner et al., 1958)

$N = 249$ Army helicopter pilot trainees in two training classes.

Median $r_{xy} = .04$ with graduation-elimination from navigator training (Melton, 1947)

$N = 1,750$ Army Air Forces navigator trainees in two training classes.

Median $r_{xy} = .14$ with graduation-elimination from bombardier training (Melton, 1947)

$N = 3,150$ Army Air Forces bombardier trainees in six training classes.

VALIDITY
(cont'd)

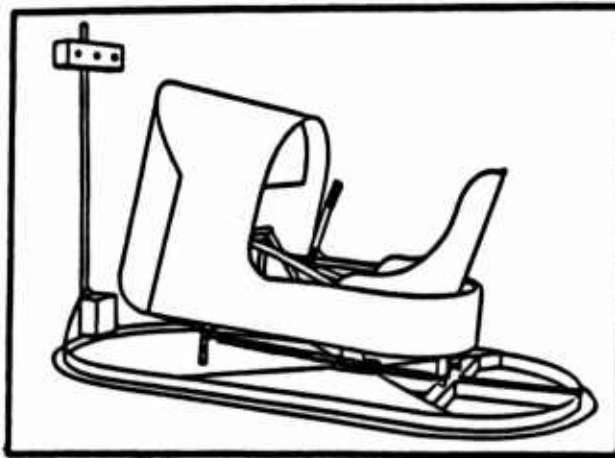
$r_{xy} = .02$ with composite flight and ground trainer grades
(Melton, 1947)

$N=47$ Army Air Forces radar operator trainees.

TEST TITLE: Rudder Control Test

**DESIGNED TO MEASURE
(CONSTRUCT):** Ability to simultaneously coordinate the movement of two feet (Multilimb Coordination)

**DESCRIPTION
OF TASK:** The subject sits in a mock airplane cockpit, which he attempts to keep lined up steadily with one of three target lights as they come on in front of him. His own weight throws the seat off balance unless he applies and maintains proper correction by means of foot pedals. The subject must also use the appropriate pedal control to shift the cockpit from one light to another as these come on at random intervals (see the figure below).



**ADMINISTRATION
AND SCORING:** The subject's score is the total time the cockpit is lined up with the proper light during the three 112-second trials. The trials are separated by 30-second rest periods.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.93$ (Melton, 1947)

Odd-even split-half reliability, corrected for the full length of the test (6 trials). $N=1,000$ unclassified candidates for pilot training.

$r_{xx}=.82$ (Fleishman, 1958)

Odd-even split-half reliability, corrected for the full length of the test (3 trials). $N=204$ basic trainee airmen.

$r_{xx}=.67$ and $.76$ (Melton, 1947)

Twenty-eight-day test-retest reliability for 311 and 312 unclassified candidates for pilot training, respectively.

VALIDITY

Median $r_{xy}=.26$ with graduation-elimination from pilot training (Melton, 1947)

$N=12,178$ Army Air Forces elementary pilot trainees in 12 training classes.

$r_{xy}=.58$ with graduation-elimination from elementary pilot training (Craeger, 1950)

$N=2,010$ Air Force elementary pilot trainees.

Median $r_{xy}=.39$ with graduation-elimination from primary pilot training (Fleishman, 1954b)

$N=4,311$ Air Force pilot trainees in two training classes.

Median $r_{xy}=.32$ with graduation-elimination from single engine jet flying training (Leiman and Friedman, 1952)

$N=3,358$ Air Force advanced pilot trainees in two training classes.

Median $r_{xy}=.46$ with graduation-elimination from pilot training (Payne et al., 1952)

$N=1,349$ Navy pilot trainees in two training classes.

VALIDITY
(cont'd)

$r_{xy} = .36$ and $.25$ with graduation-elimination from primary pilot training (Zaccaria and Cox, 1952)

$N = 1,016$ aviation cadets and 547 student officers, respectively, in primary pilot training.

Median $r_{xy} = .36$ with graduation-elimination from primary pilot training (Fleishman, 1953)

$N = 733$ Air Force primary pilot trainees in two training classes.

Median $r_{xy} = .42$ with graduation-elimination from pilot training (Mullins et al., 1968)

$N = 215$ foreign military pilot trainees in two training classes receiving training from the U. S. Air Force.

$r_{xy} = .21$ with average check ride grade (Melton, 1947)

$N = 311$ Army Air Forces pilot trainees.

Median $r_{xy} = .24$ with graduation-elimination from helicopter pilot training (Zeidner et al., 1958)

$N = 249$ Army helicopter pilot trainees in two training classes.

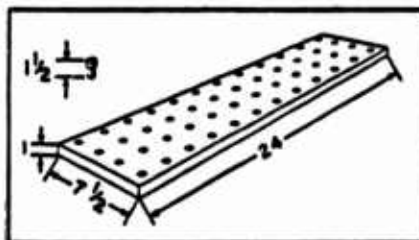
$r_{xy} = .07$ with gunner accuracy (Melton, 1947)

$N = 164$ Army Air Forces trained B-29 remote-control turret gunners.

TEST TITLE: Santa Ana Finger Dexterity Test

DESIGNED TO MEASURE
(CONSTRUCT): The ability to move and turn pegs precisely (Finger Dexterity)

DESCRIPTION
OF TASK: The test apparatus consists of 48 pegs inserted into holes in a test board (see the figure below). The holes in the board are square. The pegs themselves have square bottoms and round tops. Half of each peg top is painted blue and half is painted yellow. At the start of the test, all of the pegs are turned so that the same color of each peg top is nearest the subject. The subject's task is to pick up each peg, turn it 180 degrees so that the opposite color of the peg top is nearest the subject, and reinsert the peg into the hole.



ADMINISTRATION
AND SCORING:

The test includes five 35-second trials. Each pair of trials is separated by a 30-second rest period, during which the subject must realign his pegs so that the same color of all peg tops is nearest the subject. The subject is required to manipulate the pegs with his right hand during all five trials. The subject's score is the number of pegs turned and reinserted into the board during the five trials.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .93$ (Melton, 1947)

Odd-even split-half reliability, corrected for the full length of the test (5 trials). $N = 1,000$ unclassified candidates for pilot training.

$r_{xx} = .85, .74, \text{ and } .74$ (Melton, 1947)

Immediate, 1-week, and 28-day test-retest reliability, respectively. $N = 403, 314, \text{ and } 701$ unclassified candidates for pilot training, respectively.

VALIDITY Median $r_{xy} = .07$ with graduation-elimination from elementary pilot training (Melton, 1947)

$N = 26,032$ Army Air Forces elementary pilot trainees in 17 training classes.

$r_{xy} = .11$ with graduation-elimination from elementary pilot training (Craeger, 1957)

$N = 2,010$ Air Force elementary pilot trainees.

Median $r_{xy} = .09$ with graduation-elimination from single engine jet flying training (Leiman and Friedman, 1952)

$N = 3,538$ Air Force advanced pilot trainees in two training classes.

Median $r_{xy} = .08$ with graduation-elimination from pilot training (Payne et al., 1952)

$N = 1,368$ Navy pilot trainees in two training classes.

$r_{xy} = .11 \text{ and } .06$ with graduation-elimination from primary pilot training (Zaccaria and Cox, 1952)

$N = 1,016$ aviation cadets and 547 student officers, respectively, in primary pilot training.

$r_{xy} = .06$ with average check ride grade (Melton, 1947)

$N = 311$ Army Air Forces pilot trainees.

VALIDITY
(cont'd)

Median $r_{xy} = .14$ with instructor ratings on six pilot training performance criteria (coordination, appropriate controls, feel of controls, smoothness of movement, progress in technique, probability of success in pilot training) (Melton, 1947)

N=1,000 Army Air Forces elementary pilot trainees.

$r_{xy} = -.04$ with instructor ratings of flying proficiency (Melton, 1947)

N=562 Army Air Forces advanced pilot trainees.

Median $r_{xy} = .08$ with gunnery proficiency scores obtained during training (Melton, 1947)

N=1,716 Army Air Forces advanced pilot trainees in four training classes.

Median $r_{xy} = .14$ with graduation-elimination from navigator training (Melton, 1947)

N=2,481 Army Air Forces navigator trainees in three training classes.

Median $r_{xy} = .13$ with graduation-elimination from bombardier training (Melton, 1947)

N=4,454 Army Air Forces bombardier trainees in nine training classes.

$r_{xy} = -.32$ with percent of "on target" bombs in combat (Melton, 1947)

N=32 Army Air Forces lead bombardiers.

$r_{xy} = .22$ with gunnery accuracy (Melton, 1947)

N=164 Army Air Forces trained B-29 remote-control turret gunners.

$r_{xy} = .19$ with composite flight and ground trainer grades (Melton, 1947)

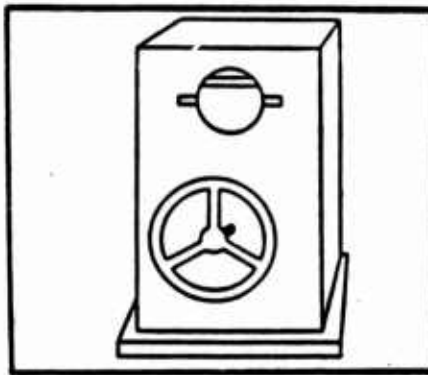
N=69 Army Air Forces radar operator trainees.

TEST TITLE: Single-Dimension Pursuitmeter

**DESIGNED TO MEASURE
(CONSTRUCT):** Rate Control

**DESCRIPTION
OF TASK:**

The subject must make compensatory adjustments (in and out movements) of a control wheel in order to keep a horizontal line in a null position as it deviates from center in an irregular fashion (see the figure below). The control wheel has been dampened pneumatically to introduce a lag into the system.



**ADMINISTRATION
AND SCORING:**

The subject receives two scores on this test. The timer score is the total time the pointer is held in a null position. The work-adder score represents the total amount of movement of the wheel during the attempt to keep the bar centered. The test consists of four 1-minute trials. The trials are separated by 15-second rest periods.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .88$ and $.92$ for the timer and work-adder scores, respectively (Melton, 1947)

Odd-even split-half reliabilities, corrected for the full length of the test (8 trials). $N=1,483$ unclassified candidates for pilot training.

$r_{xx} = .76$ for the timer score (Fleishman, 1958)

Odd-even split-half reliability, corrected for the full length of the test (4 trials). $N=204$ basic trainee airmen.

VALIDITY Median $r_{xy} = .13$ and $.06$ for the timer and work-adder scores, respectively, with graduation-elimination from pilot training (Melton, 1947)

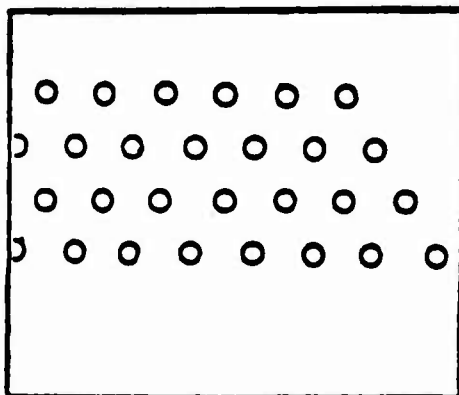
$N=836$ Army Air Forces elementary pilot trainees in two training classes.

TEST TITLE: Small Tapping Test

DESIGNED TO MEASURE
(CONSTRUCT): Aiming

DESCRIPTION
OF TASK:

The subject is presented with a piece of paper containing four rows of 10 small circles similar, but not identical, to the figure shown below. The subject's task is to place one pencil mark or dot inside each circle as rapidly as possible.



ADMINISTRATION
AND SCORING:

The subject's score is the number of circles containing a pencil mark at the end of one minute.

PSYCHOMETRIC DATA:

RELIABILITY No data are available.

VALIDITY $r_{xy} = .07$ with graduation-elimination from pilot training
(Croll, 1973)

$N=244$ South Vietnamese Air Force pilot trainees
receiving training from the U. S. Air Force.

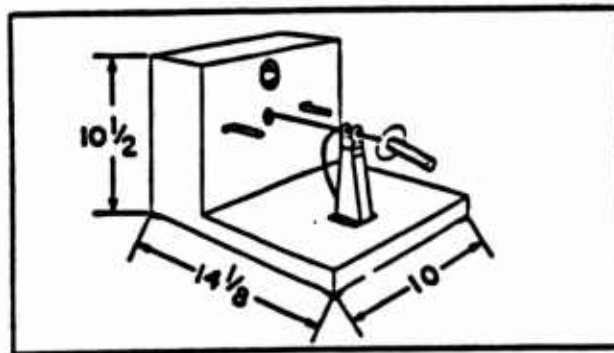
TEST TITLE: Steadiness Aiming Test

DESIGNED TO MEASURE
(CONSTRUCT):

Hand steadiness during the performance of a simple aiming task (Arm-Hand Steadiness)

DESCRIPTION
OF TASK:

The test apparatus consists of a stylus resting in a pivoted holder (see the figure below). The handle of the stylus extends down from the holder at a steep angle. The tip of the stylus is inserted inside a narrow hole or aperture. The subject's task is to hold and balance the stylus by its handle in such a way that the tip of the stylus does not touch the sides or edge of the aperture. The subject must also take care not to push or pull on the stylus, since the device has been constructed so that these movements will also cause the stylus to touch the sides of the aperture.



ADMINISTRATION
AND SCORING:

The test consists of six 40-second trials separated by 15-second rest periods. The subject receives two scores on the test. The first score is the total number of contacts between the stylus and the sides or edge of the aperture during testing, regardless of the duration of each contact. The second score is the total time the stylus spends in contact with the sides or edge of the aperture during testing.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .96$ (Melton, 1947)

Average inter-trial correlation between scores on the six trials of the test, corrected for the full length of the test. $N=461$ aviation students.

$r_{xx} = .84$ (Parker and Fleishman, 1960)

Method of estimation is unknown. $N=203$ freshman and sophomore U. S. Air Force ROTC students.

VALIDITY $r_{xy} = .13$ for total time the stylus was in contact with the sides or edge of the aperture with graduation-elimination from pilot training (Melton, 1947)

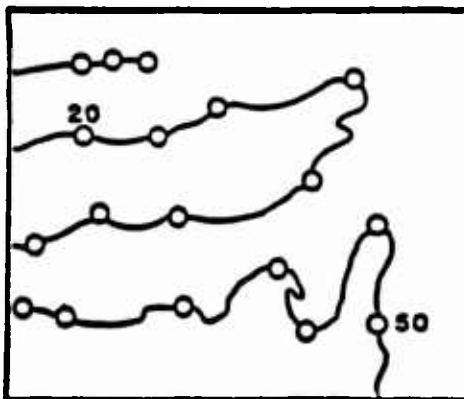
$N=516$ Army Air Forces elementary pilot trainees.

TEST TITLE: Trace Tapping II

DESIGNED TO MEASURE
(CONSTRUCT): Aiming

DESCRIPTION
OF TASK:

The subject is presented with a piece of paper containing an irregularly shaped line. One hundred small circles are placed at irregular intervals along this line (see the figure below). The subject's task is to follow the line, placing one pencil mark or dot inside each circle.



ADMINISTRATION
AND SCORING:

The subject's score is the number of circles containing pencil marks at the end of 30 seconds.

PSYCHOMETRIC DATA:

RELIABILITY No data are available.

VALIDITY $r_{xy} = .29$ with graduation-elimination from pilot training
(Croll, 1973)

$N = 244$ South Vietnamese Air Force pilot trainees
receiving training from the U. S. Air Force.

$r_{xy} = .05$ with graduation-elimination from pilot training
(Hunter and Thompson, 1978)

$N = 245$ Air Force pilot trainees.

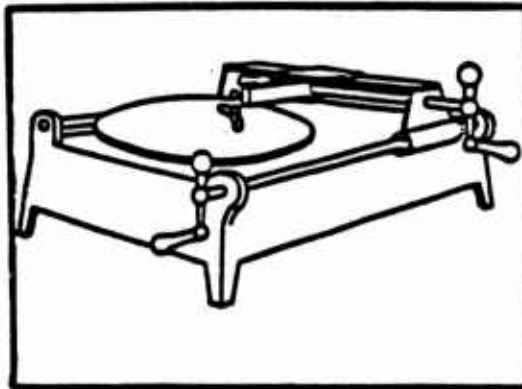
TEST TITLE: Two-Hand Coordination Test (Melton, 1947)

**DESIGNED TO MEASURE
(CONSTRUCT):**

Ability to coordinate the use of both hands to control a pursuit tracking device in response to the movements of a target which is pursuing an irregular pathway and proceeding at continuously changing rates of speed (Multilimb Coordination)

**DESCRIPTION
OF TASK:**

The target in this task is a brass button which is mounted on a phonograph-like turntable (see the figure below). The button rotates in a clockwise direction along an irregular path at varying rates of speed. The subject's task is to keep a metal leaf in continuous contact with this brass button. The position of the leaf is controlled via two rotating handles. Rotation of one handle controls forward and backward movement of the leaf, while rotation of the other handle controls side to side movement of the leaf. The two handles can be manipulated simultaneously, so that the leaf is free to move in any direction along the plane of the "turntable."



**ADMINISTRATION
AND SCORING:**

The test consists of eight 1-minute trials separated by 15-second rest periods. The subject's score on the test is the total time the leaf is in contact with the brass button (i.e., time on target).

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.90$ (Melton, 1947)

Odd-even split half reliability, corrected for the full length of the test (8 trials). $N=1,912$ unclassified candidates for pilot training.

$r_{xx}=.80$ (Fleishman, 1958)

Odd-even split half reliability, corrected for the full length of the test (4 trials). $N=204$ basic trainee airmen.

$r_{xx}=.83, .78, \text{ and } .87$ (Melton, 1947)

Immediate, one-week, and 28-day test-retest reliability, respectively. $N=416, 320, \text{ and } 700$ unclassified aircrew candidates, respectively.

VALIDITY

Median $r_{xy}=.29$ with graduation-elimination from elementary pilot training (Melton, 1947)

$N=32,260$ Army Air Forces elementary pilot trainees in 23 training classes.

$r_{xy}=.32$ with graduation-elimination from elementary pilot training (Craeger, 1957)

$N=2,010$ Air Force elementary pilot trainees.

Median $r_{xy}=.24$ with graduation-elimination from single engine jet flying training (Leiman and Friedman, 1952)

$N=3,538$ Air Force advanced pilot trainees in two training classes.

Median $r_{xy}=.38$ with graduation-elimination from pilot training (Payne et al., 1952)

$N=1,334$ Navy pilot trainees in two training classes.

$r_{xy}=.27$ and $.26$ with graduation-elimination from primary pilot training (Zaccaria and Cox, 1952)

$N=1,016$ aviation cadets and 547 student officers, respectively, in primary pilot training.

VALIDITY
(cont'd)

$r_{xy} = .21$ with average check ride grade (Melton, 1947)

$N = 311$ Army Air Forces pilot trainees.

Median $r_{xy} = .20$ with instructor ratings on six pilot training performance criteria (coordination, appropriate controls, feel of controls, smoothness of movement, progress in technique, probability of success in pilot training) (Melton, 1947)

$N = 1,000$ Army Air Forces elementary pilot trainees.

$r_{xy} = .06$ with instructor ratings of flying proficiency (Melton, 1947)

$N = 562$ Army Air Forces advanced pilot trainees.

Median $r_{xy} = .11$ with gunnery proficiency scores obtained during training (Melton, 1947)

$N = 1,193$ Army Air Forces advanced pilot trainees in three training classes.

Median $r_{xy} = .02$ and $-.09$ with three subjective ratings and four hands-on measures of flying proficiency (Lane, 1947)

$N = 37$ civilian pilot trainees.

Median $r_{xy} = .18$ with graduation-elimination from navigator training (Melton, 1947)

$N = 1,753$ Army Air Forces navigator trainees in two training classes.

Median $r_{xy} = .22$ with graduation-elimination from bombardier training (Melton, 1947)

$N = 3,531$ Army Air Forces bombardier trainees in eight training classes.

$r_{xy} = .11$ with graduation-elimination from advanced bombardier training (Melton, 1947)

$N = 423$ Army Air Forces advanced bombardier trainees.

VALIDITY
(cont'd)

$r_{xy} = .11, .14, \text{ and } .03$ with graduation-elimination from D-8 (low-altitude) bombardier training, average course grades in D-8 bombardier training, and bombing accuracy/error during D-8 bombardier training missions, respectively (Melton, 1947)

N=574 Army Air Forces D-8 bombardier trainees.

$r_{xy} = .04$ with percent of "on target" bombs in combat (Melton, 1947)

N=27 Army Air Forces lead bombardiers.

$r_{xy} = .07$ with gunnery accuracy (Melton, 1947)

N=164 Army Air Forces trained B-29 remote-control turret gunners.

Median $r_{xy} = .32$ with composite flight and ground trainer grades (Melton, 1947)

N=52 Army Air Forces radar operator trainees.

Median $r_{xy} = .15$ with supervisor ratings of overall job knowledge, overall job performance, and promotability (Helme and White, 1958)

N=1,048 Army gun crew and missile specialists, divided into nine samples on the basis of race and MOS.

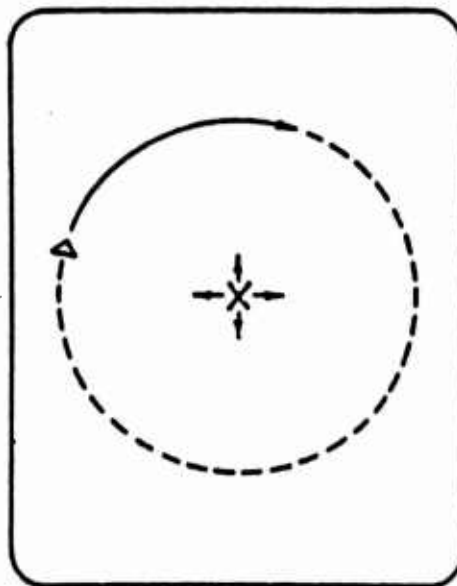
TEST TITLE: Two-Hand Coordination Test (Sanders et al., 1971)

DESIGNED TO MEASURE
(CONSTRUCT):

Ability to coordinate the movement of both hands to control a pursuit tracking device (Multilimb Coordination)

DESCRIPTION
OF TASK:

The subject uses joysticks to control the position of an X-shaped cursor displayed on a video display terminal (see the figure below). The subject's task is to track a triangle-shaped target with the cursor. The target moves around the screen in a circular path. The target's velocity changes continuously and randomly throughout the test.



ADMINISTRATION
AND SCORING:

The test consists of five 1-minute trials. There is no rest period between trials; the test is continuous. The subject receives three error scores: (1) horizontal deviation of the first stimulus from the target point, called the X Axis score; (2) vertical deviation of the first stimulus from the target point, called the Y Axis score; and (3) the square root of the sum of squares of the X Axis and Y Axis error scores (i.e., the Euclidean distance), called the Generated score.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx}=.81$ for the Generated score (Sanders et al., 1971)

Correlation between scores from minutes 4 and 5 of the test, corrected for length to represent the reliability of the sum of the two scores. $N=120$ Air Force officer trainees.

VALIDITY Median $r_{xy}=.16$, $.08$, and $.10$ with graduation-elimination from pilot training for the X Axis score, Y Axis score, and Generated score, respectively (McGrevey and Valentine, 1974)

$N=121$ Air Force officer undergraduate pilot trainees.

$r_{xy}=.19$ and $.14$ with graduation-elimination from pilot training for the X-axis score and Y-axis score, respectively (Hunter and Thompson, 1978)

$N=137$ Air Force pilot trainees.

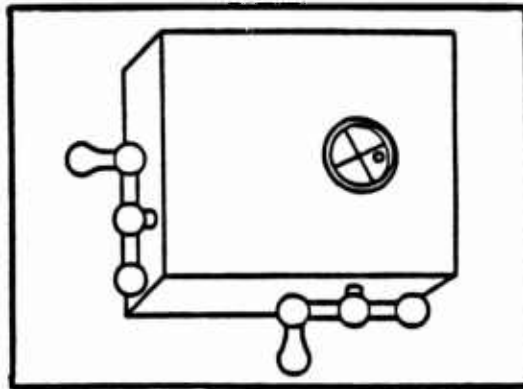
TEST TITLE: Two-Hand Pursuit Test

DESIGNED TO MEASURE
(CONSTRUCT):

Ability to coordinate the use of both hands to control a compensatory tracking device in response to the movements of a target which is pursuing an irregular pathway and proceeding at continuously changing rates of speed (Multilimb Coordination)

DESCRIPTION
OF TASK:

The target in this task is a bright piece of metal located inside a black box and superimposed against a movable black background. Both the target and the background are moving in irregular paths at varying rates of speed (see the figure below). The subject can only view the target and background through a tubular eyepiece located on the top of the box. The subject's task is to keep the target centered directly beneath a small metal button. The button is located at the intersection of a set of crosshairs. Both the button and the crosshairs are mounted at the center of the bottom of the eyepiece. The subject exerts control over the movement of the target via two rotating handles. Rotation of one handle controls backward and forward movement of the target, while rotation of the other handle controls side to side movement of the target. The two handles can be manipulated simultaneously, so that the target can be moved in any direction along the plane.



ADMINISTRATION
AND SCORING:

The test consists of eight 1-minute trials separated by 15-second rest periods. The subject's score on the test is the total time the target is centered directly beneath the metal button (i.e., time on target).

PSYCHOMETRIC DATA:

RELIABILITY Median $r_{xx} = .90$ (Melton, 1947)

Mean inter-trial correlation for the eight trials of the test, corrected for the full length of the test.

$N = 2,112$ unclassified candidates in eight aircrew training classes. The eight reliability coefficients ranged from .88 to .91.

$r_{xx} = .83$ (Fleishman, 1958)

Odd-even split-half reliability, corrected for the full length of the test (4 trials). $N = 204$ basic trainee airmen.

VALIDITY Median $r_{xy} = .26$ with graduation-elimination from pilot training (Melton, 1947)

$N = 1,169$ Army Air Forces elementary pilot trainees in eight training classes.

$r_{xy} = .20$ with graduation-elimination from bombardier training (Melton, 1947)

$N = 425$ students in bombardier training.

$r_{xy} = .20$ with graduation-elimination from advanced bombardier training (Melton, 1947)

$N = 421$ advanced bombardier trainees.

$r_{xy} = .43$ with gunnery accuracy (Melton, 1947)

$N = 32$ Army Air Forces trained Sperry turret gunners.

$r_{xy} = .43$ with circular error in estimating the lead (Melton, 1947)

$N = 32$ Army Air Forces trained Martin turret gunners.

$r_{xy} = .22$ with gunnery accuracy (Melton, 1947)

$N = 164$ Army Air Forces trained B-29 remote-control turret gunners.

VALIDITY
(cont'd)

$r_{xy} = .01$ with a composite of classroom, trainer, and flight grades (Melton, 1947)

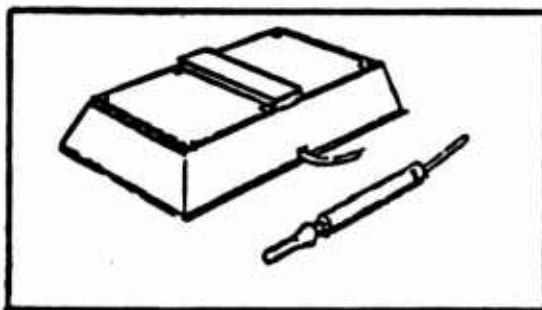
N=381 Army Air Force: radar operator trainees.

TEST TITLE: Two-Plate Tapping Test

DESIGNED TO MEASURE
(CONSTRUCT): Speed of Arm Movement

DESCRIPTION
OF TASK:

The subject must strike two adjacent metal plates with a stylus as rapidly as possible (see the figure below). He must strike the plates successively (i.e., first one plate and then the other), making as many taps as possible in the time allotted.



ADMINISTRATION
AND SCORING:

The subject's score is the number of taps made during six 30-second trials.

PSYCHOMETRIC DATA:

RELIABILITY $r_{xx} = .96$ (Melton, 1947)

Mean correlation between test scores for each of the first three minutes of the test, corrected for the full length of the test (8 minutes). $N=500$ unclassified candidates for pilot training.

$r_{xx} = .99$ (Parker and Fleishman, 1960)

Method of estimation is unknown. $N=203$ freshman and sophomore U. S. Air Force ROTC students.

VALIDITY $r_{xy} = .10$ with graduation-elimination from pilot training (Melton, 1947)

$N=1,194$ Army Air Forces elementary pilot trainees.

APPENDIX B

References to Articles, Manuals, and Technical Reports Containing Psychomotor Validity Data

**References to Articles, Manuals, and Technical Reports
Containing Psychomotor Validity Data**

- Bory, A., and Goodman, L. S. (in preparation). *Validation of a performance based pilot selection system* (NAMRL Research Report). Pensacola, FL: Naval Aerospace Medical Research Laboratory.
- Campbell, C. H., and Black, B. A. (1982). *Predicting trainability of M1 crewmen* (FR-MTRD-82-7). Ft. Knox, KY: U. S. Army Research Institute for the Behavioral and Social Sciences.
- Creager, J. A. (1957). *Validation of the February 1947 Aircrew Classification Battery for the 1950 pilot training classes* (Technical Memorandum 57-8). San Antonio: U. S. Air Force Personnel and Training Research Center.
- Croll, P. R., Mullins, C. J., and Weeks, J. L. (1973). *Validation of the Cross-Cultural Aircrew Aptitude Battery on a Vietnamese pilot trainee sample* (AFHRL-TR-73-30). Lackland Air Force Base, TX: U. S. Air Force Human Resources Laboratory, Personnel Research Division.
- de Wet, D. R. (1959). Co-ordination and floating effect. *Journal of the National Institute of Personnel Research*, 8, 28-38.
- de Wet, D. R. (1960a). A portable hand-foot reaction test. *Journal of the National Institute of Personnel Research*, 8, 106-116.
- de Wet, D. R. (1960b). An improved steadiness apparatus and its validity for air-pilot selection. *Journal of the National Institute of Personnel Research*, 8, 122-136.
- de Wet, D. R. (1962). A variable co-ordination test and its potentiality as a gauge of aptitude for airmanship. *Psychologia Africana*, 9, 86-99.
- Droege, R. C. (1968). GATB longitudinal validation study. *Journal of Counseling Psychology*, 15, 41-47.
- Eaton, N. K., Johnson, J., and Black, B. A. (1980). *Job samples as tank gunnery performance predictors* (Technical Report 473). Ft. Knox, KY: U. S. Army Research Institute for the Behavioral and Social Sciences.
- Farr, J. L., O'Leary, B. S., Pfeiffer, C. M., Goldstein, I. L., and Bartlett, C. J. (1971). *Ethnic group membership as a moderator in the prediction of job performance: An examination of some less traditional predictors* (AIR-753-9/71-TR-2). Washington, DC: American Institutes for Research.
- Fleishman, E. A. (1953a). *An evaluation of two psychomotor tests for the prediction of success in primary flying training* (Research Bulletin 53-9). San Antonio: U. S. Air Force Human Resources Research Center.

- Fleishman, E. A. (1954b). *Evaluations of psychomotor tests for pilot selection: The Direction Control and Compensatory Balance Tests* (AFPTRC-TR-54-131). Lackland Air Force Base, TX: U. S. Air Force Personnel and Training Research Center.
- Fowler, B. (1981). The Aircraft Landing Test: An information processing approach to pilot selection. *Human Factors*, 23, 129-137.
- Grant, D. L., and Bray, D. W. (1970). Validation of employment tests for telephone company installation and repair occupations. *Journal of Applied Psychology*, 54, 7-14.
- Helme, W. H., and White, R. K. (1958). *Validation of experimental aptitude tests for air defense crewmen* (PRB Technical Research Note 90). Washington, DC: Adjutant General's Office, Personnel Research Branch.
- Hunter, D. R., and Thompson, N. A. (1978). *Pilot selection system development* (AFHRL-TR-78-33). Brooks Air Force Base, TX: U. S. Air Force Human Resources Laboratory, Personnel Research Division.
- Inskip, G. C. (1971). The use of psychomotor tests to select sewing machine operators--some negative findings. *Personnel Psychology*, 24, 707-714.
- Jones, A. (1982). Providing the man in the back seat. *Proceedings of the 24th Annual Conference of the Military Testing Association*. San Antonio, TX: Military Testing Association.
- Lane, G. G. (1947). Studies in pilot selection, I: The prediction of success in learning to fly light aircraft. *Psychological Monographs*, 61, 1-17.
- Laney, A. R. (1951). Validation of employment tests for gas-appliance service personnel. *Personnel Psychology*, 4, 199-208.
- Leiman, J. M., and Friedman, G. (1952). *Validation of Aircrew Classification Battery against advanced flying training--single engine jet--criterion* (Research Note PERS 52-2). San Antonio: U. S. Air Force Human Resource Research Center.
- Loudermilk, K. M. (1966). Prediction of efficiency of lumber and paper mill employees. *Personnel Psychology*, 19, 301-310.
- Mathews, J. J., and Jensen, H. E. (1977). *Screening test battery for dental laboratory specialist course: Development and validation* (AFHRL-TR-77-53). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory, Personnel Research Division.
- McGrevey, D. F., and Valentine, L. D. (1974). *Validation of two aircrew psychomotor tests* (AFHRL-TR-74-4). Lackland Air Force Base, TX: U. S. Air Force Human Resources Laboratory, Personnel Research Division.

- Melton, A. W. (Ed.) (1947). *Apparatus tests* (Army Air Forces Aviation Psychology Program Research Report No. 4). Washington, DC: U. S. Government Printing Office.
- Mullins, C. J., Keeth, J. B., and Riederich, L. D. (1968). *Selection of foreign students for training in the United States Air Force* (AFHRL-TR-68-111). Lackland Air Force Base, TX: U. S. Air Force Human Resources Laboratory, Personnel Research Division.
- Osborn, W. C., and Ford, J. P. (1976). *Research on methods of synthetic performance testing* (HumRRO-FR-CD(L)-76-1). Alexandria, VA: Human Resources Research Organization.
- Payne, R. B., Rohles, F. H., and Cobb, B. B. (1952). *The pilot candidate selection program, IV: Test validities and intercorrelations* (Report No. 4). San Antonio, TX: U. S. Air Force School of Aviation Medicine and U. S. Navy.
- Rim, Y. (1962). The predictive validity of seven manual dexterity tests. *Psychologia*, 5, 52-55.
- Ronan, W. W. (1964). Evaluation of skilled trades performance predictors. *Educational and Psychological Measurement*, 24, 601-607.
- Shanthamani, V. S. (1978). Industrial use of dexterity tests. *Journal of Psychological Researches*, 23, 200-205.
- U. S. Department of Labor, Manpower Administration. (1970). *Manual for the USTES General Aptitude Test Battery*. Washington, DC: U. S. Government Printing Office.
- Zaccaria, L., and Cox, J. A. (1952). *Comparison of aviation cadets and student officers in primary pilot training* (Research Note PERS 52-41). San Antonio: U. S. Air Force Human Resources Research Center.
- Zeidner, J., Martinek, H., and Anderson, A. A. (1961). *Evaluation of experimental predictors for selecting Army helicopter pilots--II* (Technical Research Note 101). Washington, DC: U. S. Army TAG Research and Development Command, Human Factors Research Branch.